Efficiency Properties of Price Cap Regulation

Cloda Jenkins
Department of Economics
University College London

Thesis submitted as requirement for the Doctor of Philosophy (PhD) in Economics
Spring 2004

Primary Supervisor: Philippe Jehiel

Secondary Supervisor: Costas Meghir

Submitted for examination May 2004

Contents

List of Tables

List of Figures

Abstract

Price cap regulation has been used to control the monopoly behaviour of utility firms in the UK since 1984. The RPI-X regime currently in operation is very different to the regulatory mechanism originally proposed, and to the theory of pure price cap regulation. We therefore argue that it is necessary to examine how the mechanism is *actually* designed and implemented, if we are to determine its impact on economic welfare.

A detailed description of RPI-X regulation is presented here. This description is used to examine the impact of the price cap mechanism on allocative efficiency, technical efficiency and the efficient delivery of quality of service. We also examine the impact of regulation on welfare by calculating the rate of growth in productivity in the water and electricity sectors since privatisation.

Our analysis of the efficiency properties of RPI-X regulation exposes flaws in the regime. There is, however, scope to change the regulatory contract and potentially increase welfare. First, the way in which the price cap is set can be altered. For example, a revised methodology can be used to share cost savings with consumers. Alternatively, the form of the contract can be changed so that more restriction, than a single cap, is placed on the firm's price choice. We consider how the proposed changes may deliver a higher level of welfare given the constraints of the existing legal and institutional framework. The limited relevance and feasibility of the proposed changes, arising from the characteristics of demand and cost functions in the regulated sectors, are also emphasised.

Acknowledgements

The research for this thesis was supported by ESRC Studentship No R42200034246. I would like to thank my supervisors Prof Philippe Jehiel and Prof Costas Meghir for their advice. Helpful assistance and comments have also been provided by Jerome Adda, Heski Barr-Issac, Andrew Benito, Martin Bög, Alan Horncastle, Hide Ichimura, Chris Jenkins, Philip Lane, Tim Lane, Marco Ottaviani, Pedro Rey Biel, Richard Vaughan and Osnat Yaari. Dieter Helm, Colin Mayer and my former colleagues at OXERA provided additional support and encouragement. The librarians at Ofgem and Ofwat dealt swiftly with my many information requests, and several individuals in the regulated companies have provided me with annual reports and regulatory accounts.

I thank my family and friends for their constant support; especially my parents who have always encouraged me in my endeavours. Finally, I thank Chris for his proof-reading skills, for challenging me to look at the issues from a different perspective, and most of all for his patience and unrelenting support.

Chapter 1

Introduction

'The case for reform rests on a critique of the existing regime, and on identifying ways in which it can be improved.' (Helm, 2001)

RPI-X regulation was introduced in the UK utility sectors in 1984 when it was applied to British Telecom. Under this regulatory mechanism the rate of change in prices is limited to be equal to the change in the retail price index (RPI) minus (or plus) the expected additional net change in the firm's costs (the 'X' factor). Over the past twenty years the mechanism has been used to regulate prices in the airports, electricity distribution, electricity transmission, gas, rail, and water sectors in the UK. It has also recently been applied to National Air Traffic Control Services (NATS) and the Post Office (Consignia). This form of regulation has been adopted, and adapted, worldwide for a range of network industries in both the developed world (eg, Australia, Canada) and the developing world (eg, Latin America). It is therefore one of the most prominent forms of regulation for the utility sectors and, as such, it is important that we have a clear and complete understanding of its impact on economic welfare[1].

A cursory glance at the literature on price cap regulation generally, and on RPI-X regulation specifically, would suggest that we already know what the welfare impact is. In particular, we expect that this regulatory regime will provide the firm with an incentive to reduce costs, by allowing it to retain the profit associated with the cost reduction. This will, in theory, improve technical efficiency and thereby increase welfare. The implications for allocative efficiency and quality of service delivery are less evident in the theoretical analysis of price cap regulation.

The primary motivation for this research is the belief that the RPI-X regime, as it currently operates in the UK utility sectors, is very different from the scheme originally

[1] Armstrong, Cowan and Vickers (1994) confirm that 'Almost fifty firms in Britain are subject to RPI-X, and the system has attracted considerable international interest'. Since this book was published the price cap regime has spread even further.

proposed by Littlechild (1983), and from the theory of pure price cap regulation. For example, Littlechild (1983) envisioned a mechanism which was simple to design and implement, and which would only be required in the short-term until competition emerged. In practice, the process used to set the X-factor is complex and detailed. An analysis of operating costs, capital investment requirements, shareholder returns and output delivery is undertaken at each review. The regulator has also had to capture an increasing number of economic and political objectives in the regulatory contract[2]. In addition, price caps are expected to remain in the long-run for the natural monopoly businesses.

The theory of pure price cap regulation assumes that the price will be set independently from the firm's costs, and that the firm will retain the benefits of any savings made forever. In practice, historical and current cost information is used to set the price cap, and savings made are shared with consumers at periodic reviews (or earlier)[3].

We therefore argue that the practice of RPI-X regulation, and hence its impact on economic welfare, is different to what was originally envisioned. Given this, it is necessary to review the *actual* properties of this regulatory regime by undertaking, with the benefit of twenty years of experience with the mechanism, a thorough analysis of its impact on welfare. This analysis should be grounded in economic theory but reflect the actual characteristics of the existing regime. We therefore first describe, in **Chapter 2**, how RPI-X regulation operates in practice. The regulatory regime is presented as a series of repeated games between the firm, the regulator, and other parties. The characteristics of the RPI-X game are described, and we explain how the regulator and the firm make their decisions within this game.

The description of the game is used to analyse the efficiency properties of the price cap mechanism in **Chapter 3**. We focus on the impact of RPI-X regulation on allocative efficiency, technical efficiency and the efficient delivery of quality of service. This approach is consistent with Baumol's (1995) argument that 'the public interest standard for economic regulation calls for it to adopt only rules and procedures that are consistent with economic efficiency'. We recognise that concerns also exist about the impact of the regime on equity, but we do not address this issue here[4]. The theoretical

[2]This is emphasised by Bennett and Waddams Price (2002) who note that 'Incentive contracts must be designed to take account of several different dimensions, and have become increasingly complex'.

[3]The difference between price cap regulation in practice and in theory has been discussed in the past, particularly in the articles from the 1989 RAND Symposium on Price Cap Regulation and in Braeutigam and Panzar (1993). For example, Braeutigam and Panzar (1989) questioned 'whether any of the efficiency properties earlier ascribed to price-cap regulation will be realized'.

[4]The reader is referred to Markou and Waddams Price (1999), Waddams Price (2000b and 1997b) and Waddams Price and Hancock (1997) for an analysis of distribution concerns in the regulatory regime.

analysis is supported by an empirical investigation into the impact of regulation on productivity in **Chapter 4**. Several productivity measures are used, including changes in real unit operating costs and total factor productivity indices. We find that welfare under the mechanism is affected by the characteristics of the game, and by the way in which the X-factor is set.

The critique of the efficiency properties of RPI-X regulation leads us to our second motivation for this research. We wish to determine whether the existing regime is in fact the best available. Littlechild (1983) proposed a price cap mechanism on the grounds that it was better than the other alternatives under consideration at the time. There was no sense in which the regime was claimed to be optimal or first best. In addition, as stressed above, the system has evolved significantly in the last twenty years and, hence, it is appropriate to reconsider the question of whether it continues to be the best option.

This question is considered at two levels. First we assume, in **Chapter 3**, that it is best to retain the RPI-X mechanism and we consider ways in which the price cap setting process might be altered with the objective of increasing welfare. This analysis builds directly from the identification of a number of characteristics of the regulatory environment, and the price cap setting process, which reduce the level of welfare delivered. The feasibility of the potential improvements is also considered, taking account of the regulator's limited information set and the firm's stochastic operating environment.

We then consider a more dramatic change, in **Chapter 5**, when we assume that an alternative regulatory contract could be implemented. This analysis is based on a number of limiting assumptions, including that demand is elastic and there is a wide interval of potential cost types (i.e. cost uncertainty). It is therefore not necessarily relevant for a number of industries, such as water, which generally have inelastic demand. We compare welfare under a range of contracts which vary from no restriction on the firm's price choice, to full restriction (ie, a single price level is offered), to a single restriction (ie, the price cap), to multiple restrictions (ie, the price is constrained by the cap and one or more other restrictions). The comparisons are made under the assumption that the regulator is unable to pay a lump-sum transfer to the firm. The focus is on determining, in the first instance, whether the existing price cap contract always yields the highest level of welfare. If not, we wish to identify conditions under which a contract with more, or less, restrictions yields a higher level of welfare.

This is, by no means, the first critique of RPI-X regulation. A large volume of literature exists which examines both price cap regulation generally, and the RPI-X mechanism in particular. Individual literature references are provided on a section by section basis throughout the thesis. Our approach is similar to that adopted by

Vickers and Yarrow (1988a). They represent regulation as a repeated game between the regulator and the firm, and analyse the *expected* impact of a price cap regime in this context. The main difference is that we have the benefit of hindsight and can therefore assess how the regime actually affects decisions, and hence welfare, rather than solely relying on the predictions of economic theory. The logic of the analysis, and the economic arguments used are, otherwise, very similar.

We conclude this introduction by summarising our contribution to the literature on the economics of regulation.

- We provide a detailed description of how the RPI-X mechanism is currently designed and implemented. As far as we are aware, such a description has not been available in the academic literature before. We hope that it will provide a framework for future research into the operation and performance of this regulatory regime[5]. Our focus on the practice of regulation is in direct contrast to the theoretical literature which 'has focused on the normative rather than positive aspects of regulatory behaviour: how regulators ought to behave, rather than how they actually behave'[6].

- We present a detailed analysis of the efficiency properties of the regulatory regime. Many of the conclusions which we reach have been made by others but we bring them together in a unified framework, and we emphasise which features of the RPI-X regime are driving the efficiency properties. We combine insights from regulation economics, contract economics and information economics with our understanding of the practice of regulation[7]. We also use empirical evidence to support our theoretical analysis.

- We contribute to the literature on designing 'better' regulatory mechanisms. Several authors have focused on the institutional arrangements which need to be changed to improve welfare. This aspect has also been examined in detail by the Government in its recent review of utility regulation. We, in contrast,

[5]Laffont (1994) argued that the literature which critiqued rate of return regulation lacked 'a normative framework'. The 'new economics of regulation', which has evolved in recent decades, provides us with such a framework. The theory is, however, based on a series of limiting assumptions and, hence, the results predicted by this framework, and the proposed optimal mechanisms, can not be applied in practice. We therefore argue that Laffont's 'normative framework' needs to be amended to better reflect the actual environment within which regulation operates.

[6]Helm (1995b)

[7]Our understanding of the economic theory of optimal regulation is based on a detailed review of the research provided by Armstrong (1999), Armstrong and Sappington (2002), Armstrong and Vickers (2000), Baron (1989), Baron and Myerson (1982), Bennett and Waddams Price (2002), Caillaud, Guesnerie, Rey and Tirole (1988), Cox and Isaac (1987), Crew and Kleindorfer (2002), Gasmi, Ivaldi and Laffont (1994), Laffont (1994), Laffont and Tirole (1986 and 1993), Lewis and Sappington (1988a and 1988b), Riordan (1984), Sibley (1989) and Vickers and Yarrow (1988a).

focus on the contract itself. In particular, we present ideas on how the contract design process might be changed so as to increase welfare and we consider the feasibility of the potential improvements given the regulator's information set and the characteristics of the regulated industry. We also address the previously unasked question of whether the price cap contract imposes sufficient restriction on the firm's choice set in sectors where demand is elastic and there is significant cost uncertainty (i.e. a wide interval of potential cost types). This introduces a new dimension to the research on optimal price control contracts.

Chapter 2

The RPI-X game

When price cap regulation was first proposed for British Telecom it was put forward as a simple mechanism which would only be required in the short-term. The regulator and the firm were expected to bargain over the size of the price cap, without the need to examine inputs or outputs in any detail, and the agreed cap was to be set in the firm's licence until competition emerged in the sector[1]. The practice of RPI-X regulation has been very different to this original idea.

The regulatory regime has not 'withered away' and is expected to remain in place indefinitely for the monopoly network utility sectors. Regulators have therefore had to develop a system for changing the price cap over time. The system used to revise the cap is complicated and involves, contrary to expectations, a detailed assessment of the firm's inputs and outputs. The regulator balances the need to provide technical efficiency incentives with allocative efficiency and distributional concerns. The firm, and other parties, attempt to influence the information and judgement used by the regulator when revising the price cap. In addition, the decisions made by the regulated firm are based on an analysis of the regulator's expected actions. In this way the regulatory regime has evolved into a series of strategic interactions, or games[2]. The decisions made by the firm and the regulator, and hence the level of efficiency under the mechanism, are influenced by the playing of these games, and by the information

[1]See Beesley and Littlechild (1983) for a discussion of the expected properties of the RPI-X mechanism.

[2]Support for the idea that regulation operates as a game is provided by Bennett and Waddams Price (2002) and Vickers and Yarrow (1988a). Bishop, Kay and Mayer (1994) also argue that 'a large industry of 'regulatory games' has emerged in which firms evaluate the objectives of the regulator and optimise their behaviour conditional on the assumed response of the regulator'. Similarly the energy regulator, Ofgem (1999d) noted that 'The importance for companies of the proposals and their ability to influence the outcome in favour of their shareholders may have led to a disproportionate amount of management time and effort being devoted to management of the regulatory relationship. This, and other aspects of the application of RPI-X regulation, has led to a form of regulatory game between the regulator and the regulated companies'.

restrictions which exist.

In line with this, we describe the practice of RPI-X regulation in the UK utility sectors as a game between the firm and the regulator. Section 2.1 presents a high-level overview of the entire game. The decisions made in each stage of the game are described in Sections 2.2 and 2.3. This is a high-level, and somewhat stylized, generalisation of how RPI-X regulation works. A number of limiting assumptions are made to simplify the description, and to enable us to focus on the key features of the game which affect efficiency[3]. The description is, however, based on an in-depth review of the practice of RPI-X regulation in the utility sectors in England and Wales. It therefore captures the key features of the *actual* regulatory regime which affect welfare. We present, in Appendix A, examples of actual decisions made to justify our description of the RPI-X game.

When examining the practice of RPI-X regulation, researchers rely on the large volumes of sector-specific, and often disjointed, information produced by individual regulatory agencies. Alternatively, many academic articles simply take the price cap as a number which is given and do not explore how it is determined. This 'black box' approach means that several key features of the regime are not captured in the analysis. To overcome this gap in the literature we provide an overarching description of the way in which RPI-X regulation currently operates[4]. The game-theoretic framework allows us to bring together knowledge from the economics of game theory, information economics and regulation economics, with the practical realities of the RPI-X regime, when we analyse the welfare effects of the mechanism and when we consider ways in which the regulatory regime might be improved.

2.1 Overview of the RPI-X game

As discussed above, the RPI-X regime in the UK has evolved into a serious of repeated, and often detailed, interactions between the regulator and the firm. We present a stylised description of this RPI-X game here[5]. First we outline the moves in the

[3]A number of formulas are used to describe particular aspects of the decision-making processes. These do not necessarily represent actual formulas used in practice. Indeed, as noted by Frontier Economics (2003a), 'real regulators do not mechanically apply formulae to set price controls'.

[4]Armstrong et al (1994) provide a useful high-level description of how RPI-X regulation operates. The information is not very detailed, however, and more importantly is currently out-of-date. In particular, this book does not take account of the increased emphasis on output regulation, or the debate about how to share historical cost savings with customers.

[5]This description is an abstract summary of experiences in the water and electricity sectors between 1990 and 2000. A complementary description of how RPI-X regulation has evolved in the utility sectors can be found in Pollitt (1999). This includes an overview of the key decisions which the regulator has to make when determining both price and quality regulation. Weyman-Jones (2001b) provides a detailed description of the price-cap setting process in the 1999 electricity distribution price review.

game. Then we introduce the players. We conclude with a discussion of the main characteristics of the game which affect decision-making.

2.1.1 The two-stage game

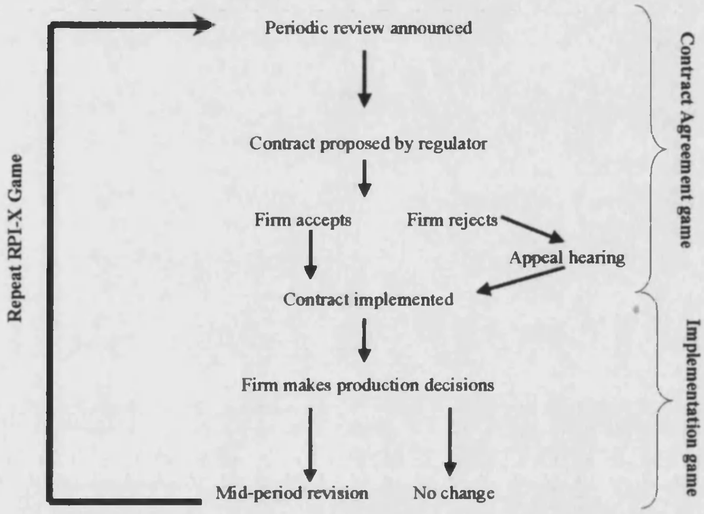

Figure 2.1: The RPI-X Game

The RPI-X Game can be classified as a repeated two-stage game. The interval between games is determined by the length of the regulatory period. The players know when each stage of the game is to be played, and when the overall game is to be repeated next. There is no known end-date to this game. The players may therefore take account of expected play in future games when making decisions.

The two stages of the RPI-X Game are illustrated in Figure 2.1 and summarised here.

1. The regulator and the firm play a **contract agreement game**[6]. The regulator offers the firm a take-it-or-leave-it regulatory contract[7]. The contract encompasses two elements - a cap on the annual rate of growth in prices, and a series

This is primarily focused on the way in which the regulator determined efficiency targets for each of the companies. Details of the regulatory regime in the gas sector can be found in Waddams Price (1997a).

[6]This game was not played at privatisation. The regulatory contract was set by the government and had to be accepted by the firm. Since then the contract agreement game has taken place at each periodic review.

[7]This may be a simplying assumption as some element of 'behind-the-scenes' bargaining no doubt takes place during the consultation period in the periodic review. This is noted by Bennett and Waddams Price (2002) who suggest that 'because of the lack of transparency, the price review is often something of a bargaining process between the regulator and the firm'. There is, however, no known *formal* bargaining process and we choose to describe the game as a 'take-it-or-leave-it' contracting

of output targets which the firm must meet for the next regulatory period[8]. If the firm accepts the contract, it is implemented in the firm's licence at the start of the next regulatory period. If the firm rejects the contract, the regulator refers the matter to an appeals body, the Competition Commission (henceforth The Commission), who issues a recommendation on how, if at all, the existing cap should be changed[9]. This alternative contract is then implemented in the firm's licence. The contract is expected to remain in place until the end of the fixed period. This stage of the game starts 12-18 months before the existing contract is due to end, and is completed approximately 3 months before the new contract comes into play. It is repeated every five-years[10].

2. The regulator and the firm play a **contract implementation game**. The firm makes its operating decisions for the regulatory period. The decision-making process is constrained by the parameters of the current regulatory contract, and by the expected impact of current decisions on the next contract agreement game. The regulator observes the firm's performance at the end of each year during the regulatory period. If the firm's decisions are different from those assumed in the contract, the regulator may choose to revise the contract early. Towards the end of the period, the regulator announces the next periodic review and a new contract agreement game begins[11]. This stage of the game is essentially on-going as it corresponds to the day-to-day operation of the regulatory firm. It is broken into five-year intervals, however, reflecting the regulatory periods to which the contracts apply.

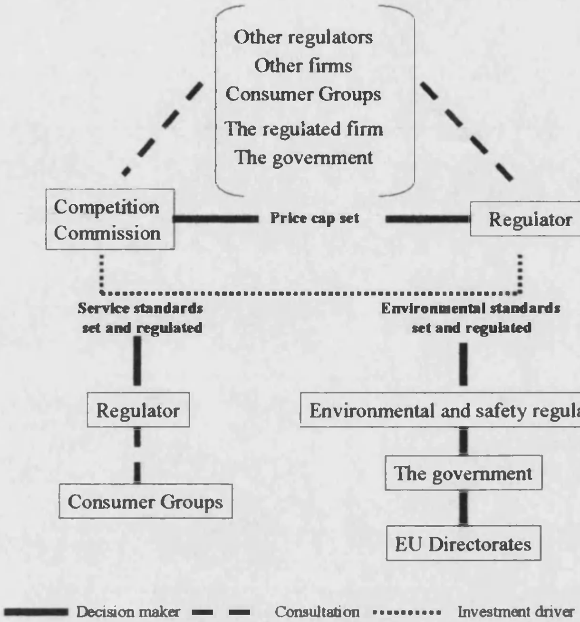

Figure 2.2: Players in the RPI-X Game

2.1.2 The Players

Figure 2.2 shows the players involved with the RPI-X game. The role of each player is discussed below. Justifications for our description of each player are presented in Appendix A. We focus on the primary decision-makers, namely the regulated firm, the economic regulator and the Commission. The other players do not make any direct decisions in the RPI-X game but their choices affect the operating environment within which the game is played.

game. We recognise, however, that the contract offer is affected by information and ideas presented by the firm, and other parties, during the consultation process.

[8]The term 'output targets' is used to describe the quality of service, environmental, and health & safety standards which the firm is required to meet on an ongoing basis. In addition, the firm must ensure that the size of its operating capacity is always sufficient to meet demand.

[9]The Competition Commission was formerly known as the Monopolies and Mergers Commission (MMC). It became the Competition Commission under the Competition Act 1998. The body responsible for utility regulation references under the Enterprise Act 2002 will be the Competition Commission Reporting Panel. We use the term 'The Commission' to cover all the names of this appeals body - past, present and future.

[10]The length of the regulatory period has varied by sector and over time. By the end of the 1990s almost all regulators had converged to a period length of five years.

[11]There is a period of overlap between the contract implementation game and the contract agreement game at the end of the regulatory period. It is unlikely that the regulator will make mid-period changes to the contract once the periodic review process begins. The firm will continue to make strategic choices for the current regulatory period. In particular, the firm is likely to make decisions at the end of the regulatory period which are focused on influencing the next contract.

- *The regulated firm:* the firm is a natural monopoly that provides a single product to consumers. It holds a licence which establishes it as the monopoly provider within a franchise area for an indefinite period. We assume that a number of regulated monopoly firms are licensed to provide the same service in other franchise areas.

  The firm is required to deliver cost savings and a range of output targets under the regulatory contract. When delivering these multiple tasks the firm's primary objective is to maximise long-run profits[12]. We assume that the firm is risk neutral and that it has a rolling two-period horizon when making decisions[13]. This reflects the idea that managers are unlikely to remain indefinitely in their jobs, and at any point in time will be most concerned about the short- to medium-term rather than the long-term. It also reflects the fact that the firm will have uncertainty about the operating environment, and the direction that regulation will take, and hence would find it difficult to make decisions for the more distant future.

  We also assume that the firm knows demand and its own costs with certainty for the ten-year horizon of interest[14]. It does not know how its decisions will affect the regulator's decisions in the next stage of the game however. Its expectations in this regard are based on what was observed in past games[15]. Expectations may be wrong ex-post because unexpected factors - such as increased pressure from the government or changes in the stock market - affect the regulator's decision. In addition, the precise form of the regulator's objective function is unknown and, hence, the firm cannot accurately forecast how decisions will be made.

- *The economic regulator:* the regulatory agency is headed by a Director General (DG) who has the final say on all decisions and is, essentially, the 'negotiating face' of the agency. The economic regulator is responsible for implementing and

[12]While profit-maximisation is the standard assumption used for a private regulated firm we note that other factors affect decision-making within the firm. Essentially a whole set of principal-agent issues arise within the firm, and between the firm and its shareholders, which may affect the firm's objectives. Weyman-Jones (2001b) emphasises, in particular, that the incentives in the RPI-X regime will be considered differently by the managers of the regulated firm and by the company owners. These alternative influences are not considered here but this is an area which warrants further research.

[13]In practice managers and shareholders are likely to be risk averse. We mention, when examining the firm's decisions and the implied efficiency properties, how risk aversion might have an influence. Our description of the game focuses on the case of risk neutrality however.

[14]In practice, demand and the firm's costs will be affected by unforeseen 'noise' during the ten-year period. As with risk aversion, we note where this uncertainty may have an effect when considering the efficiency properties of the RPI-X mechanism, but we retain the simplifying no uncertainty assumption in this description.

[15]We stress that the firm does not attempt to determine the regulator's optimal decision given the firm's choices. The decisions made may therefore not correspond to what we'd expect from the standard economic theory of strategic interactions.

revising the firm's licence and for monitoring the firm's compliance with the licence. He cannot make transfers or subsidies to the firm. In addition, the DG cannot, when setting the parameters of the regulatory contract, make any commitments beyond the current regulatory period.

The regulator is constrained by the duties set out in legislation. These require him to protect the consumer interest and to ensure that the firm is able to 'finance the proper carrying out of its functions'. We also assume that the DG is required to regulate the firm using the RPI-X mechanism[16]. The regulator has full discretion in deciding how to design the price cap itself[17].

We assume that the regulator focuses on what Mayer and Vickers (1996) call the 'Three commonly distinguished components of welfare'[18]:

(1) *Allocative efficiency* - the maximisation of welfare which results in cost-reflective pricing and the firm earning zero rent;

(2) *Technical efficiency* - the minimisation of production costs arising from the efficient use of inputs; and

(3) *Distributional concerns* - the 'curbing' of excess profits, and the desire to put economic rents in the hands of consumers rather than the firm.

The regulator considers these three objectives when making decisions and is often required to trade them off against each other. Judgement is used to decide on which of these factors is considered most important at a given point in time. This arises because, in the regulatory regime in the UK, the regulator has discretion when interpreting the duties laid out for him in legislation. More specifically, as described by Vickers and Yarrow (1988a), the regulator is required to undertake his duties 'in the manner in which he considers best calculated'. Helm and Yarrow (1988) also emphasise that the duties are 'couched in rather general terms' and, hence, there is significant scope to interpret them in different ways over time[19].

[16]In practice, regulators have the freedom to change the regulatory mechanism if they wish. This option has been discussed by a number of regulators but the RPI-X mechanism remains the regulatory tool of choice in all sectors.

[17]In contrast regulators in Latin America and the US are often constrained by detailed legal rules about how the price cap is to be designed. For example, in Chile the electricity regulator is provided with a precise set of rules about how the firm's cost information should be used to set the price cap. This is discussed in more detail in Di Tella and Dyck (2002).

[18]We purposefully don't assume that the regulator maximises a weighted expected welfare function. This assumption, paramount in the literature on optimal contract design, is considered too mechanical, and does not reflect the more discretionary approach adopted by the regulator when faced with multiple objectives.

[19]Bishop, Kay and Mayer (1995) argue that the discretion 'has created considerable uncertainty about the precise form of regulation and the way in which regulators can adjust the basis on which companies are rewarded'. We identify the specific areas in which this has occurred in Chapter 3.

The relative importance of the objectives is determined exogenously from the RPI-X Game, and may be influenced by pressures from other parties including the firm, consumer groups, financial institutions, and the Government[20],[21]. In addition, as noted by Bennett and Waddams Price (2002), the choice of a 'main goal' from amongst the set of objectives will be influenced by the regulatory structure, as well as the regulator's relationship with the government.

We assume that the regulator is only concerned about the impact of his decisions on the regulatory period in which the contract is implemented[22]. This limited foresight reflects the fact that the regulator operates on a five-year employment contract - corresponding to the regulatory period - and may not know whether that contract is to be renewed. It also reflects the constraint that the regulator is unable to make commitments beyond a single period. In addition, the short-term horizon reflects the fact that the regulator does not know the firm's cost function and is only willing to make forecasts for a relatively short period[23].

The regulator makes his forecasts using historical information about the firm and, in some cases, benchmarking analysis. The regulator's expectations are likely to be incorrect because of asymmetric information about the firm's costs.

- *The Competition Commission:* the Commission only plays in the RPI-X game if the firm rejects the regulator's proposal in the contract agreement game[24]. The Commission has the same objectives and information limitations as the regulator, but may have different expectations about the firm's decisions. It may also take a different stance on the appropriate trade-off between the three welfare objectives. The Commission considers the impact of the contract on the next regulatory period only.

[20]A richer model would be needed to capture the influence of all these interest groups.

[21]The idea that the regulator's decisions are affected by the political arena is not new, and it is not unique to the RPI-X regime. For example, Kahn (1988) argued that 'One inherent weakness of regulation is its inescapable involvement with the political process'.

[22]Regulators are required to consider the impact of all decisions on current and future customers, and on the long-run sustainability of the firm's finances. Many regulators do this by calculating the impact of the price cap on financial ratios and future price levels beyond the current period. They stress, however, that the circumstances underpinning these estimates - including the cap in the next period - are likely to be different ex-post. The primary focus of all reviews has therefore been on the impact of the contract in the next regulatory period.

[23]Jehiel (1998) provides a theoretical analysis of the implication of limited foresight on the outcome of a repeated game. In his analysis the players make correct predictions over the period of their forecast horizon, but after that the predictions are vague and random. In the case of the RPI-X game we find that other factors, including random noise and the endogenous impact of the regulatory contract, lead to a discrepancy between the forecasts and actual decisions made, even in the horizon period. The short-run rationality assumption of this paper is therefore not applicable here.

[24]Liesner (1995) presents a useful description of the role of the Commission in utility regulation for the interested reader.

- *Non-economic regulators*: the non-economic regulators operate separately from, and independent of, the economic regulator. They have two main roles in the regulated sectors: (1) they establish output targets relating to the health, safety and environmental impacts of the firm. The economic regulator must include these outputs, and their associated costs, in the final contract proposal; and (2) they are responsible for monitoring each firm's performance relative to the output targets. Where the non-economic regulator finds that the firm has breached established standards it has powers to penalise the firm with an enforcement order and/or financial penalties.

- *The government*: the government establishes the players in, and the rules of, the RPI-X game through legislation and general policy. For example, environmental policy determines the firm's output targets; the regulator's duties and powers are established in sector-specific legislation; and, the initial contract conditions were set by the government at privatisation. In addition, the government may lobby the regulator, on behalf of voters, to reduce firm profits, and thereby increase the prominence of distributional concerns from amongst the regulator's set of objectives. The government can also use taxation to share the firm's profits with consumers outside of the RPI-X game.

- *Consumers:* individual consumers are represented in the RPI-X game by national and local consumer organisations. These consumer groups play a consultative role in the regulatory game. Essentially they attempt to increase the importance of distributional concerns relative to the regulator's other objectives. The regulator has discretion about how, if at all, to reflect the consumer group's demands in his final contract decision.

2.1.3 Characteristics of the game

We conclude this overview by summarising a number of characteristics of the RPI-X Game. These factors affect decision-making and, hence, the efficiency properties of the RPI-X mechanism. The characteristics are determined by the institutional and legal framework which the firm and the regulator operate in. They are therefore considered to be a fixed element of the regulatory process. We assume that all of these characteristics are common knowledge in the game.

- The game is repeated at known fixed intervals. There is no known end-date to the game.

- The contract is revised every five years. The revised contract is proposed by the regulator but can only be implemented with the firm's agreement or after a referral to the Commission. The regulator is able to propose a contract revision at an earlier date.

- The regulator is unable to make commitments beyond the end of the regulatory period[25]. Only short-term contracts are feasible.

- We assume that the regulator regulates the firm using the RPI-X mechanism[26]. No alternative mechanisms are considered at this stage[27]. The regulator is required to protect consumer interests and to ensure that the firm earns sufficient revenues to cover the efficient costs of its operations. Beyond these legal duties the regulator has complete discretion about how to devise the regulatory contract.

- The firm faces legal action - from the regulator or in extreme cases from the Secretary of State - if it breaches the price cap condition in the contract or if it does not deliver the required level of output to consumers. Beyond this the firm has discretion about how to operate its business given the constraints of the contract.

- The regulator has multiple objectives and he makes a judgement about the appropriate trade-off across the often conflicting objectives. This judgement changes over time and may be affected by demands from the government, consumer groups, the regulated industry and other interest groups.

- The firm has a number of 'tasks' to deliver under the regulatory contract. In particular, the firm is expected to reduce costs and deliver all output targets. The incentives to deliver these multiple tasks may conflict.

- When making decisions the regulator has regard to the next regulatory period only. He knows demand for that period with certainty but does not know the firm's costs. Historical information on observed total costs of the firm, and of comparator firms, is used to form expectations.

- The firm has regard to a ten-year time horizon (two regulatory periods) when making decisions. The management know demand and costs with certainty for

[25]Laffont and Tirole (1993, Ch16) consider a number of reasons why the regulator's ability to commit might be limited by institutional arrangements.

[26]This is consistent with the practice of regulation, if not the letter of the law. The National Audit Office (2002) note that 'While there is no statutory requirement to set price controls, since privatisation the economic regulators have all used the licences to do so and have adopted a common methodology, RPI-X'. Waddams Price (2000a) also notes that 'the basic system of periodic price reviews, with price levels dependent on achieved and potential efficiency in the system, look set to stay for network parts of the industries with remaining monopoly'.

[27]In Chapter 5 we consider the impact on welfare of changing the form of the regulatory contract.

that period. At the start of the ten-year period the firm does not know what the regulatory contract will be for the second five-year period. Information about the methodology adopted at the most recent periodic review is used to form expectations about how current decisions will affect the next regulatory contract.

In many cases these characteristics, and the decisions made by the regulator and the firm, do not correspond to what we expect in standard economic analysis. In particular, expectations and decisions may not correspond with 'rational' economic behaviour[28],[29]. This is precisely why the detailed description of the game is needed. An analysis of the efficiency properties of the RPI-X mechanism, based on a presumption that the standard assumptions of economic theory hold, would not capture many of the problems which arise in practice[30].

In the next two sections we present a detailed description of decision-making in the contract agreement and implementation games. Our description of the regulator's decision-making reflects decisions observed since privatisation rather than optimal decision-making (ie, how decisions have been made rather than how we as economists think they should be made). The description of the firm's decision-making more closely resembles that expected in an economic analysis. This is because the firms do not publicise how they make their decisions and hence we are required to assume that profit-maximising behaviour is the norm.

2.2 The contract agreement game

The contract agreement game is illustrated in Figure 2.3. The main moves of the game can be summarised as follows.

1. The regulator announces the start of the periodic review about 18 months before

[28]Laffont and Tirole (1993) argue that 'Predictions about cost realizations must follow a martingale in a rational-expectations world'. In this scenario, the regulator's and the firm's expectations would always be correct. In practice, however, the expectations are frequently incorrect and we must assume that they are not rational.

[29]Gort and Wall (1988) examine the impact of the regulator's decisions on shareholder expectations, and the subsequent impact of these expectations on the decisions made. There is an assumption that all players, including the regulator and the firm, have rational expectations. We argue that this is not the case. Given this, Gort and Wall (1988) stress that there is no clear rule about how the shareholders', and presumably the firm's, expectations will differ from the actual rules used. Both over- or under-estimates can be made and other factors, including random noise, will mean that the expectations will be incorrect ex-post.

[30]Many of these characteristics are noted by Newbery (1999) who argues that 'Regulation of network utilities has to deal with asset specificity on the part of the utility, bounded rationality on the part of the regulator (incomplete and costly information about the options open to the utility), and opportunism by both parties'.

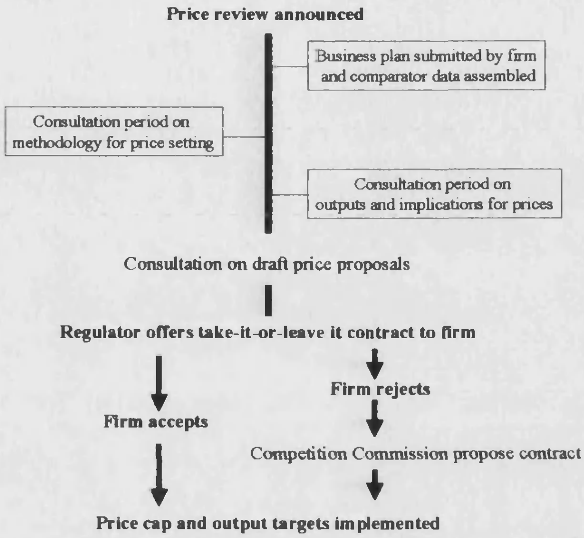

Figure 2.3: The Contract Agreement Game

the current contract is due to expire. Consultation papers on the methodology which will be used to set the price cap, and on the output targets, are published. The firm, industry groups, consumer groups, non-economic regulators, environmental organisations and other interested parties participate in this consultation process with the regulator[31]. The firm is also required to submit a business plan to the regulator. This plan incorporates the firm's forecast of its costs for the next regulatory period. No decisions are made during this consultation and information-gathering phase.

2. The regulator offers the contract to the firm. The forward-looking price cap is set so that the revenue stream under the cap is sufficient to cover expected efficient costs for the period. The output targets are set jointly with the non-economic regulators.

3. The firm decides whether or not to accept the contract offer. To make this decision, the firm compares profits under the contract to expected profits from a Commission investigation. If the firm accepts the contract, the game is over and the contract is implemented in the firm's licence.

[31]We do not present a formal description of this consultation phase here. We emphasise, however, that the 'lobbying' by different interest groups at this time can affect the regulator's decisions. There is, however, no formal process by which these viewpoints are encorporated into the final contract and hence we assume that the lobbying is an exogenous factor which affects the regulator's judgements. A richer model, incorporating the impact of this lobbying on the regulator's decision-making process, may prove useful in future research.

4. If the firm rejects the contract, a further **Competition Commission game** is played. The Commission first decides whether or not the existing cap in the licence (not that recently proposed by the regulator) is in the public interest. If it is found to be in the public interest, the cap is simply retained for the next regulatory period. If the cap is found to not be in the public interest, the Commission makes a recommendation on what changes should be made. The regulator implements a revised regulatory contract on the basis of these findings.

2.2.1 The regulator's decisions

We explain here how the regulator make his price cap and output target decisions in this stage of the RPI-X game. *The in-depth description of the contract setting process is required because it is the minute details of the contract that affect the firm's decisions and hence the efficiency properties of the RPI-X mechanism.* This is discussed in more detail in Chapter 3. The description is based on an analysis of experiences in the water and electricity sectors between 1990 and 2000. Appendix A provides a number of examples of decisions made by utility regulators in the 1990s.

The regulator uses the RPI-X mechanism to regulate the firm's price. Under this mechanism the price cannot increase, from one year to the next, by more than the rate of inflation, RPI, plus or minus a regulatory factor, X. For period j we have:

$$p^j = p^{j-1} \times \left[1 + \frac{RPI^j - X^j}{100}\right]$$

The price cap is set on a forward-looking basis and is fixed for the length of the regulatory period.

The X-factor is set so that the firm earns sufficient revenues to cover forecast efficient costs for the period[32],[33]. This approach is driven by two of the regulator's welfare objectives: first, the regulator attempts to get close to allocative efficiency by setting price equal to expected average costs; second, the regulator induces improved technical efficiency by requiring the firm to reduce costs to an efficient level, and by providing the firm with a profit-incentive, through the fixed price contract, to make additional cost savings[34].

[32]A different approach is used in US price cap schemes. Here, the X-factor is simply set equal to the historic growth rate in industry productivity, perhaps adjusted upwards by a 'stretch factor' (see Sappington, 2000). Further details can be found in Bernstein and Sappington (1999).

[33]Beesley (1994) provides an early overview of the financial modelling used to set the X-factor.

[34]As network utility industries are characterised by high sunk costs and/or increasing returns to scale, marginal cost pricing will not allow the firm to break-even. The regulator is therefore restricted to the second-best level of allocative efficiency at which price is equal to average cost.

Allowed revenues are calculated for period j as[35]:

$$\widehat{R^j} = \widehat{OC}^j + \widehat{CC}^j$$

where:

- $\widehat{R^j}$ is the allowed revenue in period j.
- $\widehat{OC}^j$ is the regulator's forecast of operating costs in period j.
- $\widehat{CC}^j$ is the regulator's forecast of capital costs in period j.

Operating costs are equal to the firm's operating productivity level, θ^j, times the amount of output sold, q^j[36]. We assume that the firm's operating effort level, e^j, is the only factor which changes the level of operating productivity[37]:

$$\theta^j = \theta^{j-1} - e^j$$

The regulator's forecast of operating costs is then calculated as:

$$\begin{aligned} \widehat{OC}^j &= \widehat{\theta}^j q^j \\ &= \left(\theta^{j-1} - \widehat{e^j}\right) q^j \end{aligned}$$

where θ^{j-1} is the level of productivity at the end of period j-1 and $\widehat{e^j}$ is the regulator's forecast of the firm's operating effort level for period j[38].

Capital costs are equal to the sum of the firm's depreciation charge and the financing cost of the asset base. The forecast depreciation charge, $\widehat{\delta}^j$, is set equal to the expected annual cost of maintaining the asset base. The financing cost of the asset base is set equal to the allowed cost of capital, $\widehat{r^j}$, times the asset value. The regulator's current cost valuation of the asset base is called the regulatory capital value (RCV). It is calculated as the value of the asset base at the start of the period - the

[35]Throughout this paper any variable with a $\widehat{(.)}$ is the regulator's forecast value for that variable. Any undecorated variable is the actual value, as chosen by the firm. Any variable with a $\overline{(.)}$ is the actual average value for the sector.

[36]We assume that the firm and the regulator know demand in the next period with certainty.

[37]In reality, operating costs will also be affected by exogenous random variables, including technological developments. The firm and the regulator will not know when these factors will arise or how they will affect the level of operating costs. Further research is required to determine how this uncertainty affects the efficiency properties of the RPI-X mechanism. We focus on the simple case of 'no noise' for our description of the game.

[38]The regulator sets the level of operating productivity at the end of period j-1 equal to the level of unit operating costs. This variable is observed at the end of each year in the firm's regulatory accounts.

opening regulatory capital value - plus the regulator's forecast of net new investment. The formula used to calculate this value is:

$$\widehat{RCV}^j = \widehat{RCV}^j_{open} + \widehat{NNI}^j(\widehat{v}^j, \widehat{s}^j, \widehat{n}^j, \widehat{\sigma}^j)$$

where:

- $\widehat{RCV}^j_{open}$ is the regulatory capital value at the start of period j;
- $\widehat{NNI}^j(\widehat{v}^j, \widehat{s}^j, \widehat{n}^j, \widehat{\sigma}^j)$ is forecast net new investment for period j[39];
- $\widehat{\sigma}^j$ is the forecast of capital productivity. This is equal to estimated productivity in the previous period minus the expected level of capital effort ($\widehat{\gamma}^j$):

$$\widehat{\sigma}^j = \widehat{\sigma}^{j-1} - \widehat{\gamma}^j$$

- $\widehat{v}^j$ is the forecast of required new capacity;
- $\widehat{s}^j$ is the quality of service target; and
- $\widehat{n}^j$ is the vector of other output targets.

The allowed revenue formula can therefore be restated as:

$$\widehat{R}^j = \widehat{\theta}^j q^j + \widehat{\delta}^j + \left(\widehat{r}^j \times \widehat{RCV}^j\right)$$

The terms in this formula are called the 'building blocks' of the allowed revenue calculation[40].

When setting operating and capital cost allowances the regulator passes on a proportion of the historical cost savings made by the firm to consumers. This results in a downward adjustment to allowed revenues. The proportion of savings which are shared, and the timing of the sharing, are determined exogenously by the regulator. These decisions are driven by distributional concerns.

As the price cap is forward-looking the regulator must make a forecast of the required costs. Here the regulator faces a problem of asymmetric information. He is not involved with the day-to-day running of the regulated firm and therefore does not

[39]When analysing the firm's costs we assume that all costs arising from quality of service, health, safety and environmental outputs are capital in nature. That is, there is no impact from these required outputs on day-to-day operating costs. In practice, the majority of investment arising from these output requirements is capital in nature but there is also an effect on operating costs.

[40]The regulator's assessment of the underlying cost requirements of the firm are used purely as a guide for determining the allowed price cap. The firm is not constrained to make cost or financing decisions which lie within the assumptions made by the regulator. It is this level of discretion which provides the firm with the incentive to look for efficient operating, investment and financing options.

know how the firm's costs evolve over time. This is a direct implication of privatisation and 'hands-off' regulation. In these circumstances, as noted by Armstrong and Sappington (2002), 'The regulator is forced to weigh the available evidence, however limited it might be, and make his best judgement about a reasonable value for the X factor'. Table A.5 in Appendix A shows the information which is used by the regulator for forecasting future cost levels. In general, the regulator uses historical information about the firm's own cost and output levels; historical information about the performance of other firms in the sector; historical information about comparator industries; and, the business plan information provided by the firms during the periodic review consultation process.

We explain in the following sections how historical savings are shared with consumers and how each allowed revenue building block is calculated by the regulator. We then describe the way in which the economic and non-economic regulators choose the output targets for the regulatory contract.

Sharing past cost savings

The price cap incorporates the regulator's forecast of the amount of operating and capital effort which the firm will undertake. Where extra cost savings have been made by the firm the regulator chooses how, if at all, to share these with consumers in the next period's price cap[41]. It is in this way that the regulator uses the periodic review to improve the distribution of economic rents.

We explain below the particular sharing rules which are used by the regulator when determining operating and capital costs for the next regulatory contract. The regulator decides on the length of the regulatory lag[42], and on whether to pass on efficiency savings in one go, or whether to have a phased sharing of savings over time. The regulator ensures, at a minimum, that the firm keeps any savings made during period j until the end of period j. There is therefore always a positive return earned from making additional cost savings. If the regulator decides to revise the regulatory contract mid-period, so as to share savings with customers before the next periodic review, the proportion of the efficiency rent retained by the firm is reduced.

[41]Under a pure price cap these efficiency rents would be retained by the firm forever. That is, the price cap would not be adjusted to reflect the difference between the regulator's expectations and the actual observed efficiency levels.

[42]This is the length of time that the firm is allowed to retain the profits from any cost saving made. It is equal to the number of remaining years in the current regulatory period, plus any additional years in the next regulatory period in which the firm is allowed to earn the efficiency rent.

Operating cost savings The regulator shares historical operating cost savings by adjusting the forecast of base operating costs[43]. Forecast base operating costs are calculated as:

$$\widehat{OC}_B^j = \theta^0 - \sum_{i=1}^{j-1} \lambda_o^{ij} e^i$$

where θ^0 is the level of productivity at privatisation; e^i is the firm's level of operating effort in period i $(i < j)$; and λ_o^{ij} is the period j sharing rule for operating savings made in period i. In this way a proportion of the reduction in operating costs is passed onto consumers over time. When $\lambda_o^{ij} = 1$ all of the savings are passed onto consumers. The length of the regulatory lag, and the decision on how sharing should be phased over time, determines the value of λ_o^{ij}.

If savings are never shared with consumers - ie, the lag is infinite - the regulator will incorporate forecast effort levels into the calculation of base operating costs giving us:

$$\lambda_o^{ij} = \frac{\widehat{e}^i}{e^i} \leq 1 \; \forall i, j$$

In contrast if all savings are shared with consumers at each periodic review - ie, the lag is equal to the length of time remaining in the regulatory period - the regulator will use actual effort levels in the calculation of base operating costs giving us[44]:

$$\lambda_o^{ij} = 1 \; \forall i, j$$

We therefore have $\frac{\widehat{e}^i}{e^i} \leq \lambda_o^{ij} \leq 1$ and can conclude that a higher value of λ_o^{ij} implies that there is more sharing of operating cost savings with consumers.

The range of actual rules used are discussed in Appendix A. In the early years after privatisation, glidepath sharing rules were used. With these rules all savings made in period j are shared gradually with consumers along a path which ensures that they are all transferred by the end of period j+1. In recent periodic reviews regulators have transferred period j savings to customers at the start of period j+1.

It is important to note that the sharing rule is determined by the regulator after the firm has made its effort choices. In addition, the regulator makes no commitment, from one review to the next, as to what the operating cost sharing rule will be and, as has been observed in practice, the rule can change over time. This is because the regulator does not use a formal rule for determining the size of the sharing rule. The

[43]Base operating costs are the required day-to-day running costs of the firm given the current level of productivity. The forecast level of operating costs is equal to the forecast level of base operating costs adjusted for expected improvement in productivity.

[44]We assume that the maximum value of the sharing rule is one. This reflects the situation in practice where profit claw-back does not occur. In principle the value could be greater than one. This would suggest that the regulator retrospectively 'claws back' the profit which the firm earns from efficiency savings and leaves the firm earning zero profits (at the extreme).

decision is based on the regulator's judgement about the appropriate distribution of rents between the firm and consumers. This judgement varies over time and is affected by many factors including whether the regulator prioritises distribution concerns ahead of technical efficiency concerns, or vice versa; whether the regulator wants to develop a reputation for being 'the consumers' champion'; whether the regulator has been captured by the regulated industry; or whether the regulator is under pressure from other interest groups, notably the government, to reduce the profits earned by the regulated firm.

Capital cost savings The regulator shares capital cost savings with consumers by adjusting the size of the opening regulatory capital value[45]. The value in period j is calculated as:

$$RCV^j_{open} = RCV^o + \sum_{i=1}^{j-1} \lambda^{ij}_c \widehat{NNI}^i$$

where RCV^o is the capital value at privatisation; $\widehat{NNI}^i$ is the forecast level of net new investment for period i ($i < j$), and λ^{ij}_c is the period j sharing rule for capital savings made in period i.

If capital savings are never shared with consumers - ie, the regulatory lag is infinite - the regulator will incorporate forecast investment levels into the calculation of the open regulatory capital value giving us:

$$\lambda^{ij}_c = 1 \ \forall i, j$$

In contrast, if all capital savings are shared with consumers at each periodic review - ie, the regulatory lag is equal to the length of the regulator period - the regulator will use actual investment levels in the calculation of the opening regulatory capital value giving us[46]:

$$\lambda^{ij}_c = \frac{NNI^i}{\widehat{NNI}^i} \ \forall i, j$$

We therefore have $\frac{NNI^i}{\widehat{NNI}^i} \leq \lambda^{ij}_c \leq 1$ and can conclude that a higher value of λ^{ij}_c implies that there is less sharing of capital cost savings with consumers. More sharing therefore implies a lower value of λ^{ij}_c but a higher value of λ^{ij}_o, because of the way that the different sharing mechanisms operate.

[45]The opening regulatory capital value is the current cost value of the firm's asset base at the start of period j.

[46]The level of actual investment considered in the sharing rule may be higher than the allowed level if additional outputs were introduced during the regulatory period. It may also, however, be lower than the amount of investment undertaken if the regulator determines that some of the firm's investment was inefficient.

In practice there has been less variation here than with operating cost sharing rules. During the 1990s regulators adopted a standardised approach of replacing allowed investment with actual investment in the regulatory capital value at the start of each regulatory period. The regulator does, however, retain the ability to change the rule from one review to the next, and any changes will reflect the same factors which influence the operating cost sharing rule.

Forecast operating costs

As noted above, the regulator determines the price cap by setting allowed revenue equal to expected efficient costs. This enables the regulator to get close to the allocative efficiency principle that prices are equal to efficient average cost[47]. We describe here how the regulator determines expected efficient operating costs and in the next section we consider expected efficient capital costs.

Forecast operating costs are set equal to efficient forecast base operating costs adjusted for expected improvements in productivity. That is:

$$\widehat{OC}^j = \widehat{OC}^j_{B^*} - \widehat{e}^j q^j$$

where $\widehat{OC}^j_{B^*}$ is the efficient level of forecast base operating costs at the start of period j, and $\widehat{e}^j$ is the expected improvement in productivity during period j.

The efficient level of forecast base operating costs is calculated as:

$$\widehat{OC}^j_{B^*} = \left[\theta^0 - \sum_{i=1}^{j-1} \lambda_o^{ij} e^i\right] - \alpha\left(OC^{j-1} - OC_*^{j-1}\right)$$

where $\alpha(.)$ is an adjustment function which reflects the difference between the firm's cost level in the previous period and the cost level of comparator firms which form the efficiency frontier $\left(OC_*^{j-1}\right)$. If the firm is deemed to have been on the efficiency frontier in period j-1 we have $\alpha(.) = 0$. The regulator uses historical information about the level of operating costs across the industry to determine the 'efficiency frontier' for the sector at the start of the period[48]. The regulator will use his discretion when deciding on the size of the adjustment.

Once the regulator has forecast the efficient level of base operating costs, he decides how these costs might be expected to change during the period[49]. The expected

[47] First-best allocative efficiency is achieved when price is set equal to marginal cost. With such a pricing rule network utility firms, which have large sunk costs, would not break even. The second-best rule of average cost pricing is therefore the only feasible option in the absence of subsidies.

[48] If costs are correlated across the firms, Frontier Economics (2003a) argue that this 'Benchmarking provides an external measure of controllable, efficient costs, in that common uncontrollable shocks are included in the benchmark but inefficiency specific to the firm is not'.

[49] A more detailed description of the methodologies used by the regulators to set efficiency targets can be found in Waddams Price (2002 and 2000a) and Jamasb and Pollitt (2000).

operating effort level is set equal to the target rate of change in the sectoral efficiency frontier ($\widehat{e}^{sector,j}$) plus an adjustment for the rate at which the firm is expected to catch-up with this frontier ($\widehat{e}^{firm,j}$). This gives us:

$$\widehat{e}^{j} = \widehat{e}^{sector,j} + \widehat{e}^{firm,j}$$

The elements of the operating effort forecast are determined as follows.

1. The sectoral element of the effort target is based on a detailed analysis of how productivity has changed over time in the sector, and how productivity has changed in other sectors in the economy in previous periods. The regulator may also examine productivity trends in the same sector in other countries to assess the potential for efficiency improvement.

2. The firm-specific element of the forecast effort level is determined by considering the firm's current efficiency level relative to the industry efficiency frontier. This relative efficiency level is assessed using econometric analysis. The firms are ranked as being above-average efficiency, average efficiency and below-average efficiency. Above-average firms are already on the industry efficiency frontier and are set a target which is similar to the expected change in the sectoral efficiency level. Average firms are close to the frontier and they are set a target that is only slightly higher than the sectoral target. Below-average firms operate well below the frontier and they are set the highest target. The high target incorporates the change needed to catch-up with other firms in the industry, and the change needed to keep up with any improvement in the sectoral frontier[50].

Appendix A provides further details of the precise methodologies adopted by the sectoral regulators.

Regulatory capital value and forecast capital investment

The regulatory capital value is calculated as:

$$\begin{aligned}\widehat{RCV}^{j} &= RCV^{j}_{open} + \widehat{NNI}^{j}(\widehat{v}^{j}, \widehat{s}^{j}, \widehat{n}^{j}, \widehat{\sigma}^{j}) \\ &= RCV^{o} + \sum_{i=1}^{j-1} \lambda_{c}^{ij} \widehat{NNI}^{i} + \widehat{NNI}^{j}(\widehat{v}^{j}, \widehat{s}^{j}, \widehat{n}^{j}, \widehat{\sigma}^{j})\end{aligned}$$

[50]For example, if the efficiency frontier is expected to shift by 100 units we have $\widehat{e}^{sector,j} = 100$. If the firm is 10 units behind the current frontier then its target would be 110 for 100% catch-up with the frontier ($\widehat{e}^{firm,j} = 10$) or 107.5 for 75% catch-up with the frontier ($\widehat{e}^{firm,j} = 7.5$). If the firm is defined as above average it may be set a target equal to 100 ($\widehat{e}^{firm,j} = 0$). Some regulators have, however, allowed 'frontier' firms a lower target than that set for the industry. For example, the regulator may set the firm an overall target of 90 units ($\widehat{e}^{firm,j} = -10$). In this way the firm is rewarded for past efficiency improvements.

The regulatory capital value at privatisation, RCV^o, is calculated using the market value of the company over the early period of trading. Adjustments may be made to this value for the specific characteristics of the sector, including differences between the market value of the firm's assets and the current accounting value.

The firm provides the regulator with a business plan for the next regulatory period during the consultation phase of the periodic review. The capital investment forecasts in this plan are used by the regulator to set forecast net new investment. A number of adjustments are made to the forecasts to ensure that they reflect what the regulator believes would be required by an efficient firm. The main adjustments made are summarised below and particular examples are given in Appendix A.

- *Efficient investment*: the regulator adjusts the firm's forecasts downwards if the proposed per unit capital costs are higher than those expected for an efficient firm at the start of period j (ie, given the current productivity level). The regulator estimates the per unit capital costs of an efficient firm by examining the historical capital cost levels of all firms in the industry, and by considering the capital costs of similar firms in other sectors.
- *Improved productivity*: the regulator compares his expected capital productivity target, $\widehat{\sigma}^j$, with the capital productivity change built into the business plan forecasts. Changes are made to the level of forecast investment to reflect any significant differences between these values. The approach used for forecasting the capital effort level is the same as that used for operating effort. In particular, the regulator uses an assessment of historical industry efficiency levels, and the firm's relative efficiency position, to set the capital effort target for period j.

The expected level of net new investment is therefore calculated as[51]:

$$\widehat{NNI}^j = NNI_f^j - \beta_1(NNI_f^j - NNI_*^{j-1}) - \beta_2(\widehat{\sigma}^j - \sigma_f^j)$$

where:

- NNI_f^j is the firm's business plan forecast of net new investment for period j;
- $\beta_1(.)$ is the adjustment for the difference between the firm's cost forecast and the expected costs of an efficient firm at the start of the period (NNI_*^{j-1}); and

[51]Regulators could also make explicit adjustments for differences between the outputs which the firm plans to deliver and those which the regulator considers are necessary. These are generally captured in the differences between the firm's investment forecast and the regulator's prediction of the required efficient level and are therefore not discussed separately here.

- $\beta_2(.)$ is the adjustment for the difference between the regulator's capital productivity target for period j and the firm's proposed improvement in productivity for the period.

The firm has an incentive to bias its forecasts upwards as this will, given the above approach to calculating the regulatory capital value, increase allowed revenues[52]. At the same time the firm knows that the downside risks from an inflated forecast are small. First, asymmetric information means that the regulator, and the independent external auditors who examine the plan, are not able to determine with any certainty whether the forecasts are honest ex-ante. Second, the firm knows that the regulator has never imposed an ex-post penalty on a firm when costs have turned out to be significantly lower than the firm's forecast. This is because the regulator does not have complete information about the firm's operations, or the technologies underlying the investment programme, and therefore cannot determine whether the firm's actual costs are lower than those forecast for legitimate reasons - eg, improved productivity - or because of overforecasting. There is thus no real financial risk to providing inflated forecasts. This means that the regulator's capital investment allowance is likely to be higher than required and, hence, ex-post inefficient.

Cost of capital

The regulator is required to ensure that the firm is able to raise the finance, from the debt and equity markets, to cover the costs of operating the business. He assumes that the firm will adopt the most efficient financing option to deliver the large-scale investment programmes. The rate of return which the firm's shareholders and investors are allowed earn is then set equal to the calculated cost of capital. Financial models, and historical data on the cost of equity and the cost of debt, are used to forecast the allowed cost of capital. The different cost of capital ranges used by the sectoral regulators are presented in Appendix A.

Depreciation

The regulator must also ensure, as part of the 'financing of functions' duty, that the firm has sufficient revenue to cover the cost of maintaining the serviceability of its

[52] Mariñoso, Hviid and Waddams Price (2002) examine, in a more theoretical framework, the impact of the regulatory regime on the firm's incentive to limit the quality of information provided to the regulator. It would be interesting to incorporate a similar model, alongside this description, in future research. This would enable us to determine the precise impact of the regulatory contract on the firm's incentive to bias its forecast, and to analyse the implication of the bias-incentive on the efficiency of the final contract.

assets in the long-run. These long-term maintenance costs are financed through an annual depreciation charge.

The allowed depreciation charge is calculated as:

$$\widehat{\delta}^j = \frac{\widehat{M}^j\left(t^j\right)}{t^j}$$

where:

- $t^j$is the average number of years over which the asset is to be maintained. The regulator uses information provided by the firm on average asset lives here; and
- $\widehat{M}^j\left(t^j\right)$ is the forecast level of efficient maintenance expenditure, given the asset life t^j. The regulator assumes that, if there is no evidence of service problems, the level of allowed maintenance expenditure in period j should be no higher than the actual amount spent in period j-1. The allowed level may be lower if the regulator believes that the firm could undertake the required maintenance work more efficiently. Efficiency is determined by comparing the firm's maintenance expenditure to its own historical levels and to the spend of other firms in the sector.

We assume that the regulator uses a straight-line methodology to determine the allowed depreciation charge. In practice, as discussed in Appendix A, this is the most common approach taken by the regulators.

Allowed revenue, service performance and the X-factor

When the forecast value of each of the building blocks has been calculated, the regulator determines allowed revenue for each year of the period by summing these components together. Before finalising the price cap, the regulator makes one further adjustment to revenue in the first year of the regulatory period. This adjustment is a financial reward/penalty for quality of service performance in the previous regulatory period. The allowed revenue calculation in the first year of period j is therefore:

$$\widehat{R}_1^j = \widehat{\theta}_1^j q_1^j + \widehat{\delta}_1^j + \left(\widehat{r}_1^j \times \widehat{RCV}_1^j\right) \pm A^j(s^{j-1}, \widetilde{s}^{j-1}, \overline{s}^{j-1})$$

where $A^j(s^{j-1}, \widetilde{s}^{j-1}, \overline{s}^{j-1})$ is the quality of service adjustment given:

- s^{j-1}, the firm's service level in period j-1,
- $\widetilde{s}^{j-1}$, the firm's service target in period j-1, and
- $\overline{s}^{j-1}$, the average quality of service in the sector in period j-1.

We describe how this adjustment operates in the water and electricity distribution sectors in Appendix A[53]. A firm is considered to have a high quality of service performance if it delivered a level of service which was greater than its target, and greater than the average level of service in the sector. A high quality firm is rewarded with an upward adjustment to the level of allowed revenue. A firm is considered to have a low quality of service performance if its level of service was below the target set. A low quality firm is penalised with a downward adjustment to the level of allowed revenue.

The regulator uses his judgement to determine the appropriate size of the adjustment[54]. He does not have the information required to base the adjustment on the actual valuation which consumers place on quality of service. Instead, the size of the adjustment reflects the regulator's confidence in the quality of the information collected on service performance; his high-level judgement about the importance of service improvement to consumers; and the belief that he should proceed with caution until the impact of the mechanism is better understood. This latter point implies that over time, as the mechanism becomes more established in the RPI-X regime, the size of the adjustment may increase.

X-factors are set which ensure that the allowed prices deliver the required net present value of revenue[55]. Several different combinations of X-factors can deliver the same net present value of revenue and the regulator uses his *judgement* as to which one is considered most appropriate. This judgement generally reflects the regulator's opinion on whether consumers would prefer price changes to be smoothed over time, or whether they prefer an immediate price reduction followed by stable or increasing prices. The regulator also considers the impact of different price profiles on the firm's financial performance by forecasting a number of key financial ratios (eg, dividend cover and interest cover) under different X-factor scenarios. In this way the regulator ensures that the chosen price profile enables the firm to finance its functions.

Output targets

The economic regulator sets standards for quality of service in the contract and incorporates the health, safety and environmental targets set by the non-economic reg-

[53]The adjustment was first used in the 1999 water and electricity distribution price reviews. The idea is not, however, new. For example, Vickers and Yarrow (1988) suggested that a quality term should be included in the X-factor formula. We presume that the proposal did not emerge in practice at that time because of a lack of information on the quality of service provided by the firms.

[54]Waddams Price, Brigham and Fitzgerald (2002) note that 'For water, the overall performance measure has a maximum of half a percent impact on revenue, with an increase likely in the future if the data improve. In electricity, up to two percent of revenue is affected in the first two years with a subsequent increase to four percent'. The authors do not explain how these revenue limits are determined however.

[55]Armstrong, Cowan and Vickers (1994, p189-192) provide details of how financial models are used to determine an appropriate range of X-factors.

ulators. The quality of service target is set to ensure, at a minimum, no deterioration in service relative to the current level. This reflects an implicit presumption that consumers always want affordable service improvements. The regulator may also decide to tighten the target. Indeed, Waddams Price, Brigham and Fitzgerald (2002) note that 'Almost all target levels have been raised since they were introduced' in the water sector and, in the electricity distribution sector, 'Required levels (ie, those triggering compensation payments) have tightened considerably across the board, particularly recently'. Appendix A provides examples of cases where regulators have increased the quality of service targets at the periodic review.

The extent to which targets are tightened is affected by the pressure placed on the regulator by the government and/or by consumer groups. The firm's productivity level will also affect the rate at which targets are changed. The regulator can choose to share historical cost savings with consumers as an improved level of quality (for a given price) rather than as a reduced price (for a given level of service). Again, this decision is made on the basis of the regulator's judgement about the appropriate price-output trade-off.

We therefore have:

$$\widehat{s}^j = \widehat{s}^{j-1} + \xi\left(s^{j-1}, s^{j-1}_{sector}\right)$$

where:

- $\widehat{s}^j$ is the quality of service target for period j;
- $\widehat{s}^{j-1}$ is the target in the previous period; and
- $\xi\left(s^{j-1}, s^{j-1}_{sector}\right)$is a possible adjustment on the previous period's target, taking account of the firm's own performance and the average performance in the sector.

The environmental and safety targets, $\widehat{n}^j$, are set by the non-economic regulators. The targets are established in primary legislation at EU and UK level. Revisions to targets are introduced when relevant legislation is changed. As these are minimum standards, required by law, there is very little flexibility in terms of deciding what needs to be delivered. The government may, however, wish to implement output targets which exceed minimum legal standards. These discretionary improvements are driven by demands for improved safety or a better natural environment. We have:

$$\widehat{n}^j = \widehat{n}^{j-1} + \zeta\left(\Delta Law, \Delta Policy\right)$$

where $\zeta\left(\Delta Law, \Delta Policy\right)$ is a tightening of the standard to reflect changes in the law and/or changes in government policy. The size of this adjustment is affected by the political party which is in power, the impact of environmental lobby groups, and the increasing importance of environmental issues at national and EU level.

We stress that consumer willingness-to-pay information has not been used to date to set the output targets. This is because it is very costly to collect reliable estimates. The regulator therefore uses his own judgement about what the appropriate price-output trade-off is. This judgement will be affected by information provided by other regulators, the industry, consumer and environmental groups, and, in some cases, scientific experts. In addition, as noted by Markou and Waddams Price (1999), the regulators have not adopted a consistent approach to setting and regulating output targets because 'the powers and obligations of the regulators are unspecified'.

2.2.2 The firm's decision

When the regulator offers the contract, the firm decides whether or not to accept it. The firm will reject the contract if it expects to earn non-positive rent - ie, if it cannot 'finance its functions'. If the profits under the contract are positive, the firm will accept the contract if profit is higher than profit under the Commission's expected contract[56]. The firm's decision is therefore[57]:

$$\begin{aligned} \text{Accept if } E(\pi_{cc}) - E(Cost_{cc}) &\leq \pi_{\mathrm{Re}g} \\ \text{Reject if } \pi_{\mathrm{Re}g} &< 0 \\ \text{Reject if } E(\pi_{cc}) - E(Cost_{cc}) &> \pi_{\mathrm{Re}g} > 0 \end{aligned}$$

where $\pi_{\mathrm{Re}g}$ is the firm's profits under the regulator's contract, $E(\pi_{cc})$ is the expected profit under the Commission's contract, and $E(Cost_{cc})$ is the expected cost of a Commission case.

We assume that the firm has all the information required to calculate $\pi_{\mathrm{Re}g}$. The firm will not be able to make a formal estimate of profits under the Commission's contract however. This is because it will be unsure of the precise values that the Commission will place on each of the 'building blocks' in the allowed revenue calculation.

The firm will instead form an expectation about whether the Commission's contract will be more or less 'generous' than the regulator's[58]. That is, the firm will calculate:

$$E(\pi_{cc}) = g(\pi^H) + (1-g)(\pi^L)$$

[56]We assume that the maximisation of expected profit is the firm's only objective. Clearly other factors, such as the management's incentive to be viewed as 'tough' with the regulator, will affect actual decision-making in this area.

[57]We assume that the firm is risk neutral. A risk averse firm would be likely to place more weight on the certain return under the regulator's contract. The management is therefore less likely to reject the regulator's contract if it risk averse.

[58]We stress that this is a high-level interpretation of how the firm *might* analyse expected profits under the Commission contract. We are unable to describe how this expectation is actually formed as there is no public information available on the firm's decision-making process.

where g is the probability that the Commission will provide a more generous contract, π^H is the expected profit from a more generous contract, and π^L is the expected profit from a less generous contract. The probability is calculated by examining previous Commission cases to determine whether the firms involved benefited from the appeals body. Expected profit under the more generous (less generous) contract is determined as a percentage increase (decrease) relative to the regulator's contract. Again this percentage adjustment will reflect the actual difference between the regulator's contract and the Commission's contract in other cases. Appendix A provides examples of recent Commission decisions, and compares them with the regulator's contract, to show the information which is available to the firm when making these calculations.

Any expected profit gain, relative to the regulator's contract, is considered net of the financial costs of a Commission investigation. These costs include legal fees, consultancy fees, the manpower and time required to put a case together, and the uncertainty surrounding the future price caps. The firm forecasts these costs by considering the in-house costs of undertaking a similar activity - for example the periodic review itself - and by analysing any published information about the costs incurred in previous cases. The firm will also consider whether the decision to reject the contract will have an impact on its future relations with the regulator. For example, it could get a reputation for being a 'trouble maker' and be treated more harshly at future reviews. This would be considered a long-term and indirect cost of the Commission investigation. Finally, the firm will consider whether a referral would have a detrimental impact on its share price, perhaps because of the uncertainty arising from a delay in agreeing a contract. If the costs are significant, the firm will be unlikely to seek a referral.

In Appendix A we find that in most periodic reviews the regulated firms have accepted the regulators' contracts. Helm (1995a) argues, in this regard, that 'Most utilities appear to have concluded that the MMC is more likely to be predisposed towards the regulator, and as a result have shied away from an MMC appeal'. In the few cases where the contract was rejected by the firm, a Competition Commission Game was played.

2.2.3 The Competition Commission Game

If the firm rejects the regulator's proposal, the case is referred to the Commission and the 'Competition Commission Game' is played. The main moves are as follows.

1. The Commission decides whether or not the current contract (ie, that set for period j-1) is in the public interest. If it is found to be in the public interest, the same contract is retained for period j and the game ends.

2. If the current contract is found to not be in the public interest, the Commission proposes an alternative contract for period j. This incorporates a price cap and details of the outputs which the firm is required to deliver.

3. The regulator designs a final contract for period j which is introduced into the firm's licence. The regulator is legally obliged to take account of the Commission's public interest finding. The regulator has discretion, however, to alter the details of the Commission's contract proposal, so long as the final contract rectifies the public interest concerns raised.

4. If the revised contract corresponds to the Commission's proposals, the firm is required to accept it. If the revised contract is different from that proposed by the Commission, the firm can seek a judicial review on the grounds that the regulator has not followed due process when devising this alternative contract. This rarely occurs[59].

The Commission's decision

The Commission's first task is to assess whether or not the existing price limit is in the public interest. There is no clear and formal definition of public interest, and we assume that the Commission uses the regulator's objectives of allocative efficiency, technical efficiency and distribution when making this judgement. If the existing contract is found to be in the public interest, the contract remains in place for period j.

The existing contract is generally found to operate against the public interest. This is because the operating environment, output targets, and the firm's costs have changed significantly since the last review and, hence, the assumptions on which the existing contract was originally based no longer apply. In this situation the Commission makes an independent assessment of what alternative contract should be introduced.

A detailed report is published outlining the assumptions used on each of the allowed revenue building blocks. The output and price cap recommendations in this report form the Commission's proposed contract for the firm. Appendix A presents a number of examples of contracts proposed by the Commission and compares them to the contracts originally proposed by the regulators. We see that the contract can be different to that proposed by the regulator. In particular, the price cap is the same, higher or lower than the one proposed by the regulator. The required output targets can also be different.

The differences arise because the Commission forms its own judgement on the appropriate value of each of the building blocks and it has access to a more recent

[59]There has only been two judicial review cases to-date: the ScottishPower case and the Northern Ireland Electricity case.

information set than the regulator. As noted above, the Commission has the same high-level objectives as the regulator. It may have a different stance, however, on the appropriate trade-off between these objectives. For example, the Commission may have an interest in being viewed as a 'consumer-friendly' organisation, and may therefore place more weight on distributional concerns.

In addition, the Commission will want to ensure consistency in its treatment of firms across the regulated sectors. Its decisions may therefore be affected by precedent set in other cases. The Commission also has a different information set from the regulator. Evidence on each building block is submitted by the regulator, the firm, the non-economic regulators and other interested parties (notably consumer groups) in the form of written reports and formal oral hearings. The economic and non-economic regulators can introduce new proposals to the Commission which were not discussed as part of the periodic review. The firm also provides an updated business plan incorporating new cost forecasts. The Commission may therefore form different expectations about the level of revenue required by the firm.

The regulator's decision

The regulator uses the Commission's published report to design a final contract for period j. If the Commission determined that the current price cap is not in the public interest, the regulator is required to change the current contract. The changes are expected to broadly reflect the Commission's recommendations, but the regulator has discretion to make amendments where he deems alternative solutions to be more reasonable. In most cases, as discussed in Appendix A, the regulator has introduced a contract which closely follows the recommendations of the Commission. The regulator's revised contract is automatically included in the licence. It can be altered at a later date, however, if the firm seeks a judicial review.

The firm's decision

The firm has very little choice over the final contract which is introduced in its licence after a Commission appeal. If the Commission decides that the current contract is in the public interest, then the firm is required to accept that the licence will not be changed. Similarly, if the regulator's final contract reflects the Commission's recommendations, the firm must accept the licence change.

If, however, the regulator chooses not to implement the Commission's contract proposal exactly, the firm can decide to take the case to the High Court for a Judicial Review. This legal review focuses on the question of whether the regulator followed due process when implementing the Commission's recommendations. Appendix A describes the two price cap judicial review cases which have been held in the utility sectors. These cases take a long time and are expensive. There is therefore a significant cost involved and hence a low probability that the firm will take this course of action.

2.3 The implementation game

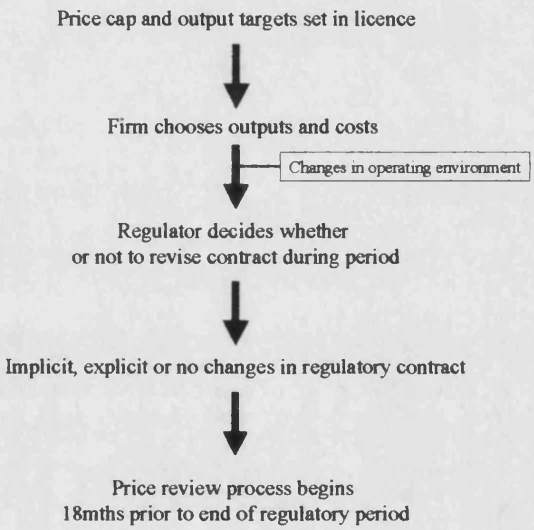

Figure 2.4: The Implementation Game

The contract implementation game is illustrated in Figure 2.4. The main moves of the game are as follows.

- The firm's licence is modified to incorporate the new regulatory contract. The contract comes into force on April 1st of the first year of the regulatory period.
- The firm makes its operating decisions for the regulatory period. These decisions are constrained by the parameters of the contract. The firm also considers the implications of its choices on the next contract agreement game.
- The economic and non-economic regulators observe the costs, profits and outputs of all firms in the sector at the end of each year during the regulatory period.
- The economic and non-economic regulators decide whether output delivery is satisfactory. Action may be taken during the period if there is evidence that the firm is not meeting its output targets.
- The economic regulator decides whether or not to revise the contract before the next periodic review.

- The regulator announces the start of the next periodic review about eighteen months before the end of the regulatory period, and a new contract agreement game begins.

2.3.1 The firm's decisions

At the start of regulatory period j the firm makes its operating decisions for the five-year period[60]. The decisions relate to the following factors: the price of the product sold (p^j); the amount of additional network capacity to produce (v^j); the amount of operating and capital effort to invest in (e^j and γ^j); the quality of service to provide (s^j); and the level of health, safety and environmental outputs to deliver (n^j)[61].

We assume, for simplicity, that the firm sets the price change to just satisfy the price cap, and that the firm produces the level of capacity required to meet long-term demand[62]. The assumption that the constraints are just binding allows us to focus on the decisions which the firm has most discretion over - namely operating and capital effort, and output delivery.

The firm maximises the future stream of expected profits given the constraints imposed by the regulatory contract. We assume that the firm is only concerned about profits in this regulatory period and in the next regulatory period. The impact of current decisions on future periods is considered negligible or too difficult for the firm to assess with any certainty. The firm therefore has a rolling ten-year time horizon. The firm is always aware, when making decisions, that there is no end-date for regulation and hence the two-period analysis is valid in all periods.

The firm's problem in period j, with a discount factor β, can be expressed as:

$$\max_{\{e^j,\gamma^j,s^j,n^j\}} E\left(\pi^j + \beta\pi^{j+1}\right)$$

subject to:

$$s^j \geq \widehat{s}^j \text{ and } n^j \geq \widehat{n}^j$$

[60]In practice, the firm is likely to have a medium-term plan for the next regulatory period and future periods. This plan will be revised from one year to the next during the period as operating conditions change. We focus on the decisions made for a single period and assume that they apply, on average, across each year of the period.

[61]We assume that operating and capital effort are independent, one-dimensional, variables. In practice, there may be many facets to each type of effort. In addition, operating and capital effort are likely to be highly correlated and substitutable. A more detailed model, along the lines of those found in the multi-tasking literature, would be required to capture the multi-dimensional nature of effort, and the interdependence across each type of effort. These issues are not considered here but do warrant further research.

[62]In practice the firm can operate below the cap and/or produce excess capacity. Markou and Waddams Price (1999) argue that 'Regulation in all the industries has been binding, suggesting that the companies would have liked to charge higher prices had the price cap permitted'. This reinforces our assumption that the firm operates at the price cap.

The constraints relate to the outputs which the firm is expected to deliver. The regulators may not be able to detect failure to deliver these outputs, particularly where lags exist between investment being undertaken and output delivery, or where monitoring is costly. In addition, the penalty for non-delivery of the output targets may not be very significant from the firm's perspective (ie, fines may be low relative to turnover). In this sense the output constraints are not 'hard', and the firm may choose to deliver output levels which are below target.

The period j profit function is:

$$\pi^j = \widehat{p}^j q^j - OC^j - CC^j - \psi(e^j) - \psi(\gamma^j) - E\left[PC^j\right]$$

where $\psi(.)$ is the cost of effort function and $E\left[PC^j\right]$ is the expected penalty cost for non-delivery of required outputs in the previous period[63]. These penalties arise in the context of statutory health, safety and environmental standards, and increase with the size of the difference between the outputs delivered, n^{j-1}, and the standard, $\widehat{n}^{j-1}$.

We assume, for simplicity, that the firm's failure to deliver a required output is observed at the end of a period and penalised in the next period. The firm will not know the exact value of the penalty. The expectation will depend on the expected probability of being caught, z, and the expected size of the penalty, ρ^j. The expected penalty cost for period j is therefore:

$$E\left[PC^j\right] = E(z\rho^j \mid n^{j-1}, \widehat{n}^{j-1})$$

The firm has the information required to calculate profits for period j. It knows demand and its cost functions for period j+1. The firm also forms an expectation about the impact of current decisions on the period j+1 price cap and output targets. Specifically, the regulator's decisions on the allowed revenue building blocks in the next periodic review are expected to mirror those at the most recent periodic review[64]. The firm does not attempt to determine the regulator's optimal reaction to its decisions, or indeed its optimal reaction to such reaction. This is because there are several factors which affect the regulator's objectives and decisions, all of which are uncertain and random from the firm's perspective. More formal and complex expectations are not formed in the face of this uncertainty.

[63]Following Laffont and Tirole (1993) we assume that the cost of effort function, and the marginal cost of effort function, are increasing and convex: $\psi'(.) > 0$; $\psi''(.) > 0$; $\psi'''(.) > 0$. In addition $\psi(0) = 0$ and $\lim_{e^j \to 1} \psi\left(e^j\right) = +\infty$.

[64]We assume that the firm has complete information about how the regulator set the elements of previous regulatory contracts (ie, the firm knows the details of the regulator's methodology, as outlined in section 2.2.1).

The operating effort decision

The price cap is set on the assumption that the firm will reduce operating costs by a set amount - $\widehat{e}^j$. At a minimum the firm ensures that it meets this target. A lower effort choice would prevent the firm from breaking-even. The firm decides whether or not to undertake a higher level of effort than this minimum requirement. A high effort choice, $e^j > \widehat{e}^j$, will improve the firm's operating productivity level and allow it to make profits. The impact on the long-run profit stream depends on how long the regulator allows the firm to have a price cap based on forecast costs which are higher than those actually incurred.

For simplicity we assume here that capital costs are zero, and we let the price cap be set such that:

$$\widehat{p}^j q^j = \widehat{OC}^j = \left(\theta^0 - \sum_{i=1}^{j-1} \lambda_o^{ij} e^i - \widehat{e}^j\right) q^j$$

The firm invests in effort level e^j, and the two-period profit stream at the start of period j is:

$$\begin{aligned} \pi^o &= \pi^j + \beta\pi^{j+1} \\ \text{where } \pi^j &= \left(e^j - \widehat{e}^j\right) q^j + q^j \sum_{i=1}^{j-1} (1 - \lambda_o^{i,j}) e^i - \psi(e^j) \end{aligned}$$

We define the firm's operating efficiency rent as the difference between profit when extra operating effort is undertaken ($e^j > \widehat{e}^j$) and profit when no extra operating effort is undertaken ($e^j = \widehat{e}^j$)[65].

$$\begin{aligned} \Delta\pi^o &= \pi\left(e^j\right) - \pi\left(\widehat{e}^j\right) \\ &= \left\{\left(e^j - \widehat{e}^j\right)\left[q^j + \beta(1 - \lambda_o^{j,j+1}) q^{j+1}\right] - \left[\psi(e^j) - \psi(\widehat{e}^j)\right]\right\} \end{aligned}$$

The firm will choose the level of effort which maximises the expected operating efficiency rent. The firm must therefore make a forecast of what the regulator's operating cost sharing rule will be at the next review. The expected sharing rule is determined by the following factors:

$$E\left(\lambda_o^{j,j+1}\right) = f\left(\lambda_o^{j-1,j}, \mu_o^j, v^j(e^j)\right)$$

where:

[65]This is similar to the approach taken by Williamson (1997), who measures the 'strength of the incentive to reduce costs' as the 'percentage of any cost saving which the regulated company can expect to keep as additional profit, *over and above what the company would earn if it did not make any saving*'.

- $\lambda_o^{j-1,j}$ is the sharing rule used to transfer savings made in period j-1 to customers in period j. The firm assumes that a similar rule will be used again at the next review. In particular, the firm will assume that the regulator will not use a rule which is more generous to the firm than the rule recently used:

$$E\left(\lambda_o^{j,j+1}\right) \geq \lambda_o^{j-1,j}$$

- μ_o^j is a correction factor which is used by the firm if the regulator has demonstrated a propensity to increase the rule from one review to the next. The firm will tighten the actual sharing rule used in the previous period by this correction factor to reflect the expectation that the regulator will introduce more sharing over time. A risk averse firm will have a larger adjustment factor as a means of insuring itself against the risk of the rule being tighter than expected. If the regulator has used similar sharing rules over time, the firm will assume that the next period's sharing rule will equal that used most recently. This gives us[66]:

$$\begin{aligned} E\left(\lambda_o^{j,j+1}\right) &= \lambda_o^{j-1,j} \text{ if } \lambda_o^{j-1,j} \simeq \lambda_o^{j-2,j-1} \\ E\left(\lambda_o^{j,j+1}\right) &= \lambda_o^{j-1,j} + \mu_o^j \text{ if } \lambda_o^{j-1,j} > \lambda_o^{j-2,j-1} \end{aligned}$$

- $v^j(e^j)$ is the expected probability of mid-period intervention by the regulator and this variable is used as a further adjustment factor by the firm. Intervention is relevant because it means that the firm has to share the benefit of any productivity improvements with customers before the next periodic review (ie, more quickly than expected)[67]. If the firm places a high probability on the regulator intervening in the next period it will increase the expected sharing rule by this factor. This gives us, potentially:

$$E\left(\lambda_o^{j,j+1}\right) = \lambda_o^{j-1,j} + \mu_o^j + v^j(e^j)$$

  The expected probability of intervention will be higher the more mid-period intervention occurred in previous periods.

The firm's problem is to choose the level of effort in period j which maximises the expected two-period operating efficiency rent . That is:

$$\max_{e^j} \left(e^j - \widehat{e}^j\right) \left[q^j + \beta(1 - E(\lambda_o^{j,j+1}))q^{j+1}\right] - \left[\psi(e^j) - \psi(\widehat{e}^j)\right]$$

[66]We assume that the firm only considers experiences in the previous two periods.

[67]Mayer and Vickers (1996) stress that mid-period intervention means that the regulatory lag, and hence the sharing rule, is not fixed exogenously. That is, the lag is affected by the regulator's decision to intervene and hence 'Unless credibility not to intervene is total, the effective lag, so to speak, will be shorter than the explicit lag'.

The first-order condition of this problem gives us the familiar condition that the firm chooses effort up to the point where the marginal benefit of effort equals the marginal cost[68]:

$$q^j + \beta(1-)q^{j+1} = \psi'(e^j)$$

This decision rule reflects the direct impact of the firm's effort decision on its long-run profits. The higher the expected level of sharing - i.e. the higher $E(\lambda_o^{j,j+1})$ - the lower the level of savings made, and the higher the level of output in period j the higher the level of effort undertaken.

The firm also considers the impact of the current operating effort choice on other elements of the next period's price cap. The firm does not know exactly how the price cap will be set next period, although it may know the direction of the relationship between the cap and other variables, and hence it must form an expectation of the size of the change. The size of this effect, $\frac{\partial E(\widehat{p}^{j+1})}{\partial e^j}$, will depend on the following factors.

- The firm's operating effort level increases its productivity level and, hence, increases the productivity of the industry on average. The higher the average industry improvement, the higher will be the sectoral operating effort target in the next period.

- A high effort choice in period j may improve the firm's operating cost level relative to other firms in the sector. In this case, the regulator will deem it to be 'efficient' at the next periodic review and it will be set a less stringent effort target for period j+1, relative to other firms who have to catch-up with the efficiency frontier.

- A high effort choice in period j will reduce the firm's operating costs and will make it difficult for the firm to propose a higher level of costs in the business plan submitted at the next periodic review.

Taking these direct and indirect effects into account, the firm's operating effort decision is to set the expected marginal benefit of increasing effort equal to the marginal cost. The optimisation condition will be[69]:

$$q^j + \beta(1 - E(\lambda_o^{j,j+1}))q^{j+1} - \psi'(e^j) \pm \beta\left(\frac{\partial E(\widehat{p}^{j+1})}{\partial e^j}\right) = 0$$

[68]The second-order condition is $-\psi''(e)$ which ensures a local maximum with our convex cost of effort function. This is the second-order condition for all our profit-maximisation and welfare-maximisation problems.

[69]We assume that an improvement in productivity can lead to either a reduction or an increase in the price limit.

The firm's expected sharing rule may turn out to be significantly different from that which is used by the regulator. This is because other factors, such as changes in the way that the regulator priorities the technical efficiency and distribution objectives, and/or increased pressure from the government for reduced prices, affect the regulator's decision. These factors are not known to the firm when making its effort choice - ie, they are 'noise' in the expectation - and hence are not captured in the expected sharing rule.

The capital effort decision

The firm's decision with respect to capital effort is very similar to that outlined above for operating effort. In particular, the firm has a profit-incentive to choose a higher level of effort than assumed by the regulator, and that level of effort will be chosen by reference to a marginal benefit equal to marginal cost condition.

We assume here that there are no operating costs, and the price cap is set so that:

$$\begin{aligned} \widehat{p}^j q^j &= \widehat{r}^j \times \widehat{RCV}^j \\ &= \widehat{r}^j \times \left[RCV^o + \sum_{i=1}^{j-1} \lambda_c^{ij} \widehat{NNI}^i + \widehat{NNI}^j \right] \end{aligned}$$

To enable us to focus on the effort decision we assume that the difference in capital effort is the only factor which causes the allowed capital level to be different from the actual capital level. We also assume that the cost of capital assumption is correct. This gives us:

$$\begin{aligned} \widehat{r}^j &= r^j \\ \widehat{NNI}^j - NNI^j &= \gamma^j - \widehat{\gamma}^j \end{aligned}$$

The firm invests in effort level γ^j and the relevant profit stream from period j forward is:

$$\pi^c = \pi^j + \beta \pi^{j+1}$$

$$\pi^j = r^j \times \left(\sum_{i=1}^{j-1} \left(\lambda_c^{ij} \widehat{NNI}^i - NNI^i \right) + \left(\gamma^j - \widehat{\gamma}^j \right) \right) - \psi(\gamma^j)$$

The firm's capital efficiency rent from undertaking a higher than assumed level of capital effort capital ($\gamma^j > \widehat{\gamma}^j$) is:

$$\begin{aligned} \Delta \pi^c &= \pi \left(\gamma^j \right) - \pi \left(\widehat{\gamma}^j \right) \\ &= \left(\gamma^j - \widehat{\gamma}^j \right) \left(r^j + \beta r^{j+1} \lambda_c^{j,j+1} \right) + \psi(\widehat{\gamma}^j) - \psi(\gamma^j) \end{aligned}$$

The firm will form an expectation of the capital sharing rule at the next periodic review. As with the expected operating sharing rule, this will be an adjusted function of the sharing rule used at the previous price review[70]:

$$\begin{aligned} E\left(\lambda_c^{j,j+1}\right) &= \lambda_c^{j-1,j} \text{ if } \lambda_c^{j-1,j} \simeq \lambda_c^{j-2,j-1} \\ E\left(\lambda_c^{j,j+1}\right) &= \lambda_c^{j-1,j} - \mu_c^j \text{ if } \lambda_c^{j-1,j} > \lambda_c^{j-2,j-1} \end{aligned}$$

A downward adjustment, μ_c^j, is made to the previous period's rule if the regulator demonstrated a propensity to sharing more capital savings with consumers (i.e. reduce $\lambda_c^{j,j+1}$) from one period to the next. We stress, again, that the sharing rule used in the next periodic review may be very different from the firm's expectation.

The firm's problem is to choose the level of effort in period j which maximises the expected profit stream. That is:

$$\max_{\gamma^j} \left(\gamma^j - \widehat{\gamma}^j\right)\left(r^j + \beta r^{j+1} E\left(\lambda_c^{j,j+1}\right)\right) + \psi(\widehat{\gamma}^j) - \psi(\gamma^j)$$

The first-order condition of this problem gives us:

$$r^j + \beta r^{j+1} E\left(\lambda_c^{j,j+1}\right) = \psi'(\gamma^j)$$

That is, the marginal benefit from an extra unit of capital effort is set equal to the marginal cost of an extra unit of capital effort. The higher the expected capital sharing rule - i.e. the lower the expected level of sharing of capital savings - the higher the level of effort. Similarly, the higher the cost of capital the higher the level of effort.

The firm must also consider the impact of increased capital effort on the next period's price cap, $\frac{\partial E(\widehat{p}^{j+1})}{\partial \gamma^j}$. The main effects on the price cap are the same as in the case of operating effort[71]. The firm's capital effort decision is therefore determined by the condition that:

$$r^j + \beta r^{j+1} E\left(\lambda_c^{j,j+1}\right) - \psi'(\gamma^j) \pm \beta\left(\frac{\partial E(\widehat{p}^{j+1})}{\partial \gamma^j}\right) = 0$$

The capital investment decision

The firm will undertake capital investment if it expects to be able to finance that investment for the duration of the capital project. The regulator sets the allowed level

[70] The regulator is unlikely to intervene mid-period to require the firm to share capital savings with customers before the next periodic review. This is because low capital costs at the start of a period may simply reflect an alternative profiling of the investment programme and higher costs may emerge later in the period. The regulator can therefore not attribute short-run capital cost reductions to a sustained improvement in productivity.

[71] The reader can pinpoint the effects by simply replacing the word 'operating' in our prior discussion with the word 'capital'.

of investment at the forecast efficient level. This may be less than the level which the firm believes is required for the project. In addition, the regulator only provides a guaranteed rate of return for the duration of the regulatory period. The firm is therefore left with some uncertainty as to whether it will earn sufficient revenues in future periods to offset the actual financing costs of the on-going investment project.

The firm's decision with regard to a T-period capital investment project can be summarised as follows.

1. If the regulator has not allowed sufficient financing for the investment project in period j - the start of the project - the firm will not undertake the investment:

$$\text{No investment if } r^j NNI^j \geq \widehat{r}^j \widehat{NNI}^j$$

2. If the regulator allows sufficient revenues for the project in period j, the firm will consider the difference between expected required and allowed financing in future periods. The project will not be undertaken if the allowed financing is lower than required expected financing. This will occur if the firm believes that the regulator intends to 'hold-up' the firm at a later date, either by reducing the allowed level of investment or by reducing the allowed cost of capital. If the project has begun, and this happens, the firm will be required to self-finance the remainder of the project itself. To avoid this risk, the firm will choose not to undertake the project in the first place:

$$\text{No investment if } E\left(\sum_{i=j+1}^{T} r^i NNI^i\right) > E\left(\sum_{i=j+1}^{T} \widehat{r}^i \widehat{NNI}^i\right)$$

   The firm is also aware that only investments which are allowed under the regulatory contract will be financed by the regulator. Other discretionary projects will therefore not be undertaken.

3. The firm will undertake the investment project if it expects the regulator to provide adequate financing for the duration of the project. This will occur if the firm expects that the regulator will pay due regard to his duty to ensure it can 'finance its functions' in all periods.

$$\text{Investment if } E\left(\sum_{i=j+1}^{T} r^i NNI^i\right) \leq E\left(\sum_{i=j+1}^{T} \widehat{r}^i \widehat{NNI}^i\right)$$

A similar situation arises with maintenance expenditure. The allowed depreciation charge is determined over a longer length of time than the regulatory period, and the regulator may change the method of calculating that charge from one period to the next. Say the asset has an expected life of T-years and the efficient cost of maintaining it for that period is M. If the regulator sets the depreciation charge such that $\widetilde{\delta}^j < \frac{M}{T}$, the firm will be reluctant to undertake the required expenditure. The firm will only be willing to undertake maintenance of the assets if it is sure that the depreciation charge is sufficient to cover the long-run costs involved.

The quality of service decision

We noted above that the output constraints which the firm faces may not always be satisfied. The firm may choose to deliver an output level which is below the target set if there are financial benefits to doing so. We assume that when the firm reduces investment, by delivering a lower level of output, the impact on customers or the environment is immediate (ie, in the current year), and is observable and verifiable by the regulator[72],[73].

We define the net benefit of delivering quality of service level s^j as:

$$NB(s^j) = \Delta\pi\left(s^j\right) + H\left(s^j\right) - C\left(s^j\right) \pm \beta E\left(A^j(s^{j-1}, \widetilde{s}^{j-1}, \overline{s}^{j-1})\right)$$

where:

- $\Delta\pi\left(s^j\right)$ is the profit earned from choosing this level of service. This value is zero if the firm simply produces the regulator's target, assuming all of the regulator's other assumptions are correct.
- $H\left(s^j\right)$ is the intrinsic reputation value which the firm earns from the quality of service provided. This value function is increasing in the quality of service provided[74].

[72] In practice, a firm may be able to reduce investment without there being an immediate effect on the output delivered because the network condition is sufficient to provide the required outputs in the short-term. The investment reduction will, however, jeopardise the long-term serviceability of the network and increase the risk of service problems in the future. This adds a further dimension to the firm's decision - trading off the risk of an output delivery problem being observed, and hence punished, in the current period against the potential of it not being observed for some time. This extra element of the decision problem is not considered here but warrants further analysis.

[73] Dalen (1997) and Laffont and Tirole (1993, Ch4) consider the optimal regulatory contract when quality is unverifiable.

[74] Laffont and Tirole (1993) argue that reputation incentives are important if 'The firm's main incentive to provide quality is...the threat of jeopardizing future trading opportunities with the buyer rather than current ones'. In the regulated utility sectors the firm has a long-term licence to provide services in the franchise area and consumers do not have the option of switching to another supplier if the quality of service provided is insufficient. In these long-term monopoly industries, therefore, the reputation effect may not be significant. The regulator could attempt to increase the size of the reputation effect by introducing a licence condition which allowed the regulator to revoke the licence, and issue it to another firm, if the current licence-holder underdelivered with respect to quality of service. The idea of placing restrictions on the licence length is one which warrants further attention from the regulators.

- $C\left(s^j\right)$ is the level of compensation which is paid directly to consumers affected by service problems. This is zero if the quality of service is at or above the target set.

- $E\left(A^j(s^{j-1}, \hat{s}^{j-1}, \bar{s}^{j-1})\right)$ is the expected size of the quality of service adjustment made to the next period's price cap. The expected size of the service performance adjustment will be correlated with the size of the adjustment used at the last review (A^{j-1}).

To enable us to focus on the quality of service decision, we assume that the difference between the service level and the service target is the only factor which causes the assumed investment level to be different from the actual investment level. In particular we assume that all other required output targets are delivered, and the regulator's assumptions on expected productivity and the cost of capital are correct. This gives us:

$$\begin{aligned}
\widehat{r^j} &= r^j \\
\widehat{NNI}^j - NNI^j &= \widehat{s^j} - s^j
\end{aligned}$$

The profit stream from period j forward is then:

$$\begin{aligned}
\pi &= \pi^j + \pi^{j+1} \\
\pi^j &= r^j \times \left(\sum_{i=1}^{j-1} \left(E(\lambda_c^{ij})\widehat{NNI}^i - NNI^i \right) + \left(\widehat{s^j} - s^j \right) \right)
\end{aligned}$$

The rent earned by the firm, when it produces quality of service level s^j, is the difference between the profit stream when this quality of service level is undertaken and the profit stream when the quality of service target is undertaken ($s^j = \widehat{s^j}$).

$$\begin{aligned}
\Delta\pi\left(s^j\right) &= \pi\left(s^j\right) - \pi\left(\widehat{s^j}\right) \\
&= \left(\widehat{s^j} - s^j\right)\left(r^j + \beta r^{j+1} E(\lambda_c^{j,j+1})\right)
\end{aligned}$$

If the firm delivers the target standard of service it will earn zero profits. If the firm delivers a higher quality of service than the target set, s_h^j, it will make a loss. If the firm delivers a quality of service that is below the target, s_l^j, it will make a profit. Profits will be higher the lower the level of expected capital cost sharing (i.e. the higher is $E(\lambda_c^{j,j+1})$).

The firm will produce the level of service which yields the highest net benefit. The net benefit of delivering $\widehat{s^j}$ is:

$$NB(\widehat{s^j}) = H\left(\widehat{s^j}\right) + \beta E\left(A\left(\widehat{p}^{j+1} \mid A\left(\widehat{p^j}\right), \widehat{s^j}, \bar{s}^j\right)\right)$$

The net benefit of delivering a higher level of service, s_h^j, is:

$$NB(s_h^j) = -\Delta\pi\left(s_h^j\right) + H\left(s_h^j\right) + \beta E\left(A\left(\widehat{p}^{j+1} \mid A\left(\widehat{p}^j\right), s_h^j, \widehat{s}^j, \overline{s}^j\right)\right)$$

The net benefit of delivering a lower level of service, s_l^j, is:

$$NB(s_l^j) = \Delta\pi\left(s_l^j\right) + H\left(s_l^j\right) - C\left(s_l^j\right) - \beta E\left(A\left(\widehat{p}^{j+1} \mid A\left(\widehat{p}^j\right), s_l^j, \widehat{s}^j, \overline{s}^j\right)\right)$$

The firm must also consider the indirect impact of service performance in period j on the period j+1 contract. The level of service delivered may impact on the price cap in the following ways.

- If a low level of service is delivered, the regulator will require the firm to catch up with the target without allowing additional investment for this on the grounds that the financing was already provided by customers.
- A high level of service will signal to the regulator that a higher service target is feasible for period j+1. This may be applied at industry level or only to the firm itself. The high target is likely to be set at current cost levels, requiring the firm to deliver more with the same level of revenues.
- A low level of service, and the associated reduction in capital investment costs, will make any attempt by the firm to argue for high investment costs in its next business plan incredible to the regulator.

The 'other outputs' decision

The firm's decision-making process with respect to health, safety and environmental standards is very similar to that outlined above for quality of service. The main difference is that there is no formal adjustment in the price-cap setting process[75]. Instead, the firm faces a lump-sum penalty from the non-economic regulators. The absence of any reward structure for outperforming the target means that the firm only considers whether to meet the target or breach it.

We define the net benefit of delivering output level n^j as:

$$NB(n^j) = \Delta\pi\left(n^j\right) + M\left(n^j\right) - \beta E(z\rho^{j+1} \mid n^j, \widehat{n}^j)$$

where:

[75]The service index used to assess the water companies' performance at the 1999 review incorporated variables relating to both quality of service and environmental standards. We assume here that the price cap adjustment only reflects quality of service performance (as was the case with electricity distribution) and that separate penalties exist for the other output targets. This ensures that the firm is not penalised twice for non-delivery of these outputs.

- $\Delta\pi\left(n^j\right)$ is the profit earned from choosing this output level. This value is zero if the firm simply produces the regulator's target and all of the regulator's other assumptions are correct.
- $M\left(n^j\right)$ is the intrinsic reputation value which the firm earns from the health, safety and environmental outputs provided. This value function is increasing in the level of output provided.
- $E(z\rho^{j+1} \mid n_l^j, \widehat{n}^j)$ is the expected penalty which the firm faces from the non-economic regulators in period j+1. The firm only faces this penalty if it produces a lower output level than the target, n_l^j.

Following exactly the same argument as we used for quality of service, we define the firm's profit from choosing output level n^j as:

$$\Delta\pi\left(n^j\right) = \left(\widehat{n}^j - n^j\right)\left(r^j + \beta r^{j+1}\lambda_c^{j,j+1}\right)$$

Again, the firm chooses the output level which yields the highest net benefit. The net benefit of delivering $\widehat{n}^j$ is:

$$NB(\widehat{n}^j) = M\left(\widehat{n}^j\right)$$

The net benefit of delivering a lower output level , n_l^j, is:

$$NB(n_l^j) = \Delta\pi\left(n_l^j\right) + M\left(n_l^j\right) - \beta E(z\rho^{j+1} \mid n_l^j, \widehat{n}^j)$$

The firm will only choose the target level of output if the expected penalty cost is sufficiently high. The firm must also consider the impact of output delivery in period j on prices and profits in period j+1. The relevant factors are the same as those outlined above for quality of service delivery. In addition, the probability of being punished for non-delivery of required outputs will be higher in period j+1 if the firm was caught breaching the target in period j. This is because the non-economic regulators will increase the level of monitoring on poorly performing firms. The size of the penalty will also increase if the firm breaches the target repeatedly.

2.3.2 The regulator's decision

The regulator observes the firm's actual cost levels, and outputs delivered, at the end of each year of the regulatory period. The regulator is therefore able to see whether his cost forecasts were too high or too low, and whether the firm is delivering the required output targets. In addition, the regulator sees whether there have been changes in the firm's operating conditions which may affect its ability to finance its functions.

When this information becomes available the regulator has two decisions to make in the implementation game. First, the regulator decides whether the firm needs to be penalised for underperforming with respect to output targets. Second, the regulator decides whether the regulatory contract should be revised mid-period in light of the information now available.

Output delivery

The regulator monitors the firm's quality of service performance each year. An annual report is published showing how the firm performed relative to its target, and relative to other firms in the sector. This 'name and shame' approach is intended to provide a reputation incentive to the firm to meet its quality of service target.

If the firm fails to meet a quality of service target it must pay compensation to those customers who are affected. If there is a serious breach of a service target, or if there are persistent and repeated problems, the regulator can issue an enforcement order. This is a legal instrument which requires the firm to take action to rectify the identified problem without the provision of financial assistance. The economic regulator does not impose any other direct penalties on the firm during the regulatory period but, as discussed in section 2.2.1, an adjustment will be made to the next period's price cap to reflect service performance.

The non-economic regulators monitor the firm's performance with respect to the other output targets. As with quality of service, the firm's performance is published in an annual report. If the firm breaches a target, the non-economic regulator has the legal power to impose a penalty on the firm. The probability of a penalty being imposed depends on the extent of the breach and the level of monitoring undertaken by the regulator. The non-economic regulator can also use enforcement orders to change the firm's behaviour.

The non-economic regulators may introduce revised output targets during the regulatory period. These changes are driven by new legislation. The impact on the firm's profits will depend on how the economic regulator chooses to treat the associated cost increase. As discussed below, the regulator may choose to revise the contract mid-period and allow for the increased costs to be financed through the price cap. Alternatively, the regulator may decide that the firm can self-finance the additional output delivery for the regulatory period, and will include the investment in the regulatory capital value at the next periodic review.

Mid-period contract revision

During the period the regulator may observe that the assumptions underlying the regulatory contract are very different to what actually emerges. If this is the case, the regulator decides whether or not to revise the contract mid-period so that it reflects the updated information. In this way the incompleteness of the contract - ie, the fact that it does not cover all contingencies - makes commitment to the contract for the fixed period difficult.

There is no set principle or objective which the regulator uses when choosing whether or not to intervene. We present, in Appendix A, a number of cases where mid-period intervention occurred. The regulator's decision depends on how he prioritises his three welfare objectives. The main benefits to revising the contract are that it improves allocative efficiency by making the price reflective of recent actual cost information, and it distributes efficiency rents from the firm to consumers earlier than expected. The cost is that the firm earns a lower share of any cost savings made, and therefore has a reduced incentive to improve technical efficiency. In future periods, the firm will take account of the mid-period intervention when forming its expectation of the sharing rule.

We note that the contract may also be changed in the firm's favour. If output targets are tightened, the regulator may increase the price cap to cover the extra investment costs. This is most evident in the water sector, where provisions exist for the licence to be changed when the non-economic regulators introduce new output requirements during the regulatory period.

2.4 Concluding remarks

We have presented a detailed description of the way in which the regulator and the firm make their decisions in the RPI-X game. In addition, a number of fixed characteristics of the current regime have been highlighted. The description reflects the experiences which have been observed in the water and electricity sectors since privatisation. In Chapter 3 we use this description to analyse the impact of the RPI-X regime on allocative and technical efficiency and on the efficient delivery of outputs. We also examine ways in which the details of the regime, and the mechanism itself, might be changed so as to improve its efficiency properties.

Chapter 3

Welfare under the RPI-X regime

Littlechild (1983, 1986), and others, argued that price cap regulation was the best available means of regulating the profitability of utility companies because it provides the firm with an incentive to minimise costs[1]. This technical efficiency advantage is also predicted by the economic theory of pure price cap regulation. The theory is based on the assumptions that the cap is set independent of the firm's costs, and that the firm retains all of the rent from cost savings made. We know, from our description in Chapter 2, that the RPI-X mechanism does not operate in this way. Historical information is used to set cost forecasts, and the regulator shares a proportion of the efficiency rent with consumers over time. The practice of price cap regulation therefore differs from the theory. This must be explicitly recognised if the true welfare effects of the RPI-X mechanism are to be accurately determined. In addition, cost minimisation is only a partial measure of welfare improvement. The net welfare effect can only be assessed by also examining the impact on allocative efficiency and on the efficient delivery of outputs[2].

Bearing these points in mind, we consider the welfare properties of the RPI-X mechanism in section 3.1. Specifically, we examine the impact of the price cap regime on allocative efficiency, technical efficiency and the efficient delivery of outputs[3]. The analysis is used to determine whether the RPI-X mechanism is delivering the maximum feasible level of welfare for a fixed price cap[4]. The analysis is based on our description

[1]The expected benefits of price cap regulation are discussed in Beesley and Littlechild (1983) and Vickers and Yarrow (1988a).

[2]This argument was also made by Vickers and Yarrow (1988a) who stress that 'a finding that private firms have lower unit costs than their public counterpart does not necessarily imply that their contributions to social welfare are greater; questions relating to allocative efficiency and to the quality of goods or services provided also need to be taken into account'.

[3]We do not consider the impact of the regulatory regime on the distribution of benefits. A discussion of this equity issue can be found in Markou and Waddams Price (1999).

[4]A large number of articles have been written which compare the price cap mechanism to alternative mechanisms, including rate of return regulation and sliding-scale regulation. The interested reader is

of the RPI-X regime in Chapter 2 and, hence, is clearly grounded in the practice of price cap regulation rather than the high-level theory of pure price caps.

Our objectives our two-fold: (1) to enable the reader to better understand the welfare properties of RPI-X regulation; and, (2) to focus attention on the particular features of the RPI-X game which affect these welfare properties. We present, in section 3.2, a number of alternative regulatory mechanisms which build on the existing RPI-X regime. The alternatives involve changes in the current price cap setting process. We assume that the game characteristics are fixed by the institutional and legal framework of utility regulation in the UK, and will affect any regulatory mechanism which is used[5]. We explain how these alternative mechanisms *may* lead to an improvement in allocative efficiency, technical efficiency and/or the delivery of an efficient level of quality of service. We also discuss the feasibility of the potential improvements being delivered given the limited information which is available to the regulator and the stochastic nature of demand and hence costs in the regulated network utility sectors.

Many of the arguments made here have been made by others, both practitioners in the regulated sectors and commentators[6,7]. These studies are generally discursive and high-level in nature. They highlight a number of concerns with the price cap mechanism, but they do not examine the specific elements of the RPI-X regime which affect the efficiency properties of this form of regulation. In addition, many of the papers are now quite dated and do not reflect the way in which the mechanism has evolved in recent years. We contribute to the debate on the merits of RPI-X regulation by bringing the existing arguments together in one place. We also ground the analysis in a framework which enables us to better understand how the nature of demand and costs in the industry, the characteristics of the RPI-X game, and the regulator's

referred to the following papers, and the references therein, for further details: Armstrong, Cowan and Vickers (1994), Bennett and Waddams Price (2002), Braeutigam and Panzar (1993), Gasmi, Ivaldi and Laffont (1994), Gilbert and Newbery (1994), Helm and Yarrow (1988), Lewis and Sappington (1989), Liston (1993), Lyon (1996), Mayer and Vickers (1996), Pint (1992), Sappington (2000), Schmalansee (1989), Viehoff (1995), Waterson (1992 and 1995), Weisman (1993 and 1994) and Yarrow (1989).

[5]Bennett and Waddams Price (2002) provide a complementary discussion of the institutional and political factors which affect the power of incentive mechanisms.

[6]Early critiques of the RPI-X regulatory regime can be found in Armstrong, Cowan and Vickers (1994), Beesley and Littlechild (1983, 1989), Glynn (1992), Helm (1995a and 1995b), Helm and Yarrow (1988), Rees and Vickers (1995), Vickers and Yarrow (1988a and 1988b), Waterson (1992 and1995) and Yarrow (1989). More recent analyses of the regime, which reflect changes made at the first and second rounds of periodic reviews, are provided by Bennett and Waddams Price (2002), Bernstein and Sappington (1999), Cowan (1997), Crew and Kleindorfer (1996), Green and Haskel (2001), Helm (2001), Littlechild (2001), Lowe (1998), Markou and Waddams Price (1999), Mayer (2000), National Audit Office (2002), Newbery (1999) and Waddams Price (1997a).

[7]Critiques of a broader range of regulatory incentive mechanisms, including pure price caps, are provided by Crew and Kleindorfer (1996), Gasmi, Ivaldi and Laffont (1994), Laffont and Tirole (1993 and 2000), Parker (2002), Sappington (2000), Vogelsang (2002), and the papers in the 1989 RAND Symposium on price cap regulation.

decision-making process affect the efficiency properties of price cap regulation. Any proposed changes must take account of these features to ensure that the potential for welfare improvement is linked to our understanding of the actual limitations of the existing mechanism given the characteristics of the game and the regulated industry.

3.1 The welfare properties

In this section we consider the impact of the RPI-X mechanism on the three elements of welfare. Section 3.1.1 focuses on allocative efficiency, technical efficiency is examined in section 3.1.2, and we examine the impact of RPI-X regulation on efficient output delivery in section 3.1.3.

Conflicts arise in the RPI-X game which prevent all types of efficiency from being delivered simultaneously. It is therefore difficult to measure the overall welfare impact of this mechanism. We find that technical efficiency incentives are, at least partially, offset by concerns about allocative efficiency and the provision of required output targets. In addition, the cost minimisation incentives are dulled by the way in which the RPI-X game is played. The primary benefit of price cap regulation may therefore not be as strong in practice as envisioned by the theory of pure price caps.

3.1.1 Allocative efficiency

We consider here whether the regulatory contract delivers allocative efficiency in the RPI-X game. Allocative efficiency is achieved when welfare is maximised given current costs. When a firm has large fixed costs and/or increasing returns to scale - as in most network utility companies - marginal cost pricing is not feasible and we focus on average cost pricing as the appropriate benchmark. The firm earns zero profits with these prices.

Figure 3.1 shows the average rates of return earned in the water and electricity sectors since privatisation[8]. These suggest that allocative efficiency has not been delivered as firms are earning above normal returns[9]. We explain below why this problem arises with the RPI-X mechanism.

Under the allowed revenue approach, discussed in section 2.2.1, the price cap is essentially set equal to expected average efficient cost in the RPI-X game.

$$\widehat{p}^j = E(AC_*^j)$$

[8]The rate of return is measured as current cost operating profit divided by the regulatory capital value. The data on current cost operating profit is taken from the firms' annual regulatory accounts. The data on the regulatory capital value is taken from the regulatory agencies's publications.

[9]As shown in Table A.14, the allowed pre-tax weighted average cost of capital has been set at less than 8% by all regulators. The actual rate of return has exceeded this benchmark in all sectors and in all years since privatisation.

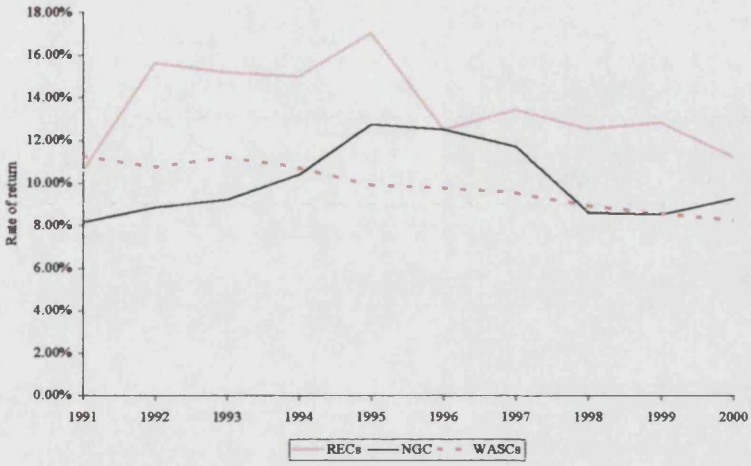

Figure 3.1: Rates of return in the regulated sectors

We therefore have allocative efficiency if the regulator's expectation of average efficient cost is equal to the actual level of average cost. There are several reasons why this will not be the case.

1. Asymmetric information about the firm's underlying technology, and the unobservability of effort, mean that there is a positive probability that the regulator's forecast of expected costs will be inaccurate. This is a problem which exists for all regulatory mechanisms.

2. In the utility sectors both demand and costs are stochastic in nature. In particular uncertain exogenous factors, such as weather and other random events (e.g. a World Cup football game leading to peak demand for electricity), affect demand levels, capacity utilisation and hence required costs. The regulator therefore has uncertainty about how the firm's operating environment will change during the regulatory period and there is noise in the regulator's forecast[10]. Again, this is not unique to price cap regulation.

3. The regulator uses the firm's forecast of capital costs when setting allowed revenues. The firm has an incentive to inflate this forecast. The expected capital costs used to set the price cap are therefore likely to be higher than required. This widens the gap between actual and expected average cost.

[10]Schmalansee (1989) argues that in this situation a welfare-maximising regulator will set a higher allowed price cap to provide insurance to the firm, assuming the firm is risk averse. This moves the firm further away from an allocatively efficient price level.

The fixed nature of the price cap exacerbates the allocative efficiency problem further. Even if the regulator's expected average costs are correct at the start of the period the ex-post costs will, as a direct result of the technical efficiency incentives, be lower than the current costs[11]. There is thus a divergence between the allowed price level and average costs at the end of the period[12]. In this way the existence of multiple regulatory objectives means that attempts to deliver one objective prevent another from being achieved[13]. This is similar to the results of the multi-tasking literature, epitomised by Holmstrom and Milgrom (1991).

Several authors have found that the welfare-maximising contract requires the regulator to pay a positive transfer to the firm[14]. In these analyses, the option of paying a zero transfer is available but is never found to yield the highest level of welfare. Under current legislation, the regulator is unable to pay transfers to the firm. We can therefore conclude that welfare is below the maximum level because transfers are restricted.

Finally, in an attempt to improve distribution, the regulator shares efficiency savings with customers over time. The sharing rules are chosen arbitrarily by the regulator. In particular, they are not chosen to maximise long-run welfare. Indeed, Mayer (2000) stresses that the choice of the five-year regulatory lag 'has more to do with regulatory and political expediency than economic theory'. The final price level may therefore be distorted further away from the constrained welfare maximising level because the sharing rules are not optimal.

We conclude that allocative efficiency is not delivered under the RPI-X mechanism. The features of the RPI-X game which affect this element of welfare can be summarised as follows.

1. *Asymmetric information and uncertainty* - the regulator can only achieve second-best welfare levels when he is uncertain about the firm's future cost and demand

[11]Mayer (1999) emphasises that the system 'hinges crucially on projections that are inevitably disproved almost as soon as they are made'.

[12]A similar problem occurs with other optimal incentive mechanisms where the regulator is required to pay a rent to the firm to induce it to undertake the optimal level of effort. These mechanisms are discussed in detail in Laffont and Tirole (1993).

[13]It should be noted that the allocative efficiency problems of the RPI-X mechanism were recognised from the outset. A decision was made that the delivery of productive efficiency today was more important than allocative efficiency. This short-run trade-off reflected the assumption that, particularly with respect to British Telecom, competition would emerge to reduce the extent of allocative efficiency. This competition has not emerged for the natural monopoly network firms, and regulator's have used the periodic review process as a means of reducing the allocative efficiency problem.

[14]For example, Laffont and Tirole (1986) find, in their model of regulation with moral hazard and adverse selection, that the optimal contract involves the firm setting price equal to marginal cost and the regulator paying the firm a transfer to induce it to decrease its costs. There is therefore allocative efficiency, but at the expense of some technical efficiency. In the absence of these transfers, this optimal contract can not be implemented and there is some loss of allocative efficiency and technical efficiency.

conditions. There is always a positive probability that allowed average costs will differ from actual average costs.

2. *Restricted transfers* - the regulator does not have the power, or the finances, to pay lump-sum transfers to the firm. Optimal regulatory mechanisms, proposed in the literature on the new economics of regulation, are therefore not feasible[15].

3. *Multiple objectives* - the regulator allows the firm to earn a rent to provide technical efficiency incentives, but this immediately results in allocative inefficiency. Attempts to deliver one objective therefore reduce the regulator's ability to deliver another.

4. *Regulatory discretion* - the sharing rules for operating and capital savings are chosen arbitrarily. We therefore expect a deviation from welfare maximisation as the price levels, incorporating these sharing rules, are not set so as to maximise long-run welfare.

Characteristics (1) to (3) are ingrained in the RPI-X game and will prevent any regulatory mechanism from delivering allocative efficiency. The fourth feature is unique to the RPI-X mechanism, as currently designed, and could *in theory* be changed by the regulator. However, as discussed in section 3.2.1, asymmetric information and uncertainty about future demand and costs will restrict the regulator's ability to determine a welfare maximising sharing rule in practice. Attempts could also be made to alleviate the allocative inefficiency problem by improving the regulator's information set. We explore these potential improvements further in section 3.2.

3.1.2 Technical efficiency

Price cap regulation was introduced for the utility sectors in England and Wales because it was thought to provide the firm with an incentive to reduce costs. We show, in Chapter 4, that RPI-X regulation has delivered the expected cost savings. There is a question, however, as to whether the cost savings are as large as they could be, and whether they are sustainable in the long-term. We also question whether the cost saving incentives ensure that the firm chooses the most efficient combination of inputs.

The main problems with the technical efficiency incentives are summarised here and discussed in detail below.

1. The regulator shares cost savings with customers. This reduces the firm's incentive to make optimal effort choices.

[15]See Laffont and Tirole (1993).

2. The sharing rules change over time. The firm is therefore uncertain about the actual benefit it will earn when it makes its effort choices. This may dull the effort incentives further.

3. The regulator may adjust the sharing rule in response to changes in the size of the cost savings made. If this occurs there is a ratchet effect problem. The firm's incentive to reduce costs is offset by the knowledge that lower costs will trigger a higher sharing rule and, hence, reduce the gain from each unit of effort[16].

4. If the regulator revises the contract mid-period, the firm's expected sharing rule may be increased and future effort levels will fall.

5. Historical information is used to determine future cost allowances. This introduces an additional potential ratchet effect problem. The firm weighs up the efficiency rent from reduced costs against the reduction in future profits from prices being based on lower cost levels.

6. The incentive regime is different for operating and capital cost savings. It also varies across years in the regulatory period. This distorts the firm's input choices and thereby reduces technical efficiency.

The incentive to reduce costs

In Chapter 4 we provide evidence on the extent to which firms have reduced their operating and capital costs since privatisation. Real unit operating costs have reduced significantly since privatisation in the electricity distribution, electricity transmission and water sectors. In contrast, real unit capital costs have increased although the increases *may* have been lower than would have been the case in the absence of RPI-X regulation. There has also been an increase in total factor productivity, but the average annual growth rates are relatively small.

We explained, in section 2.3.1, how the firm is provided with the incentive to reduce costs. In the words of Vickers and Yarrow (1988a), we argue that the incentives are in fact more 'illusory' than 'real'. Our reasons are discussed below.

Impact of sharing rules

The firm has an incentive, under a fixed *pure* price cap, to undertake the welfare-maximising level of effort. This is because it retains all of the rent earned from reducing

[16]Papers which analyse the ratchet effect in the regulatory environment include Armstrong and Sappington (2002), Currie, Levine and Rickman (1999), Dalen (1997 and 2000), Freixas, Guesnerie and Tirole (1985), Laffont and Tirole (1988 and 1993, Chs 9&10) and Lewis and Sappington (1997). A more general analysis of the ratchet effect is presented in Bolton and Dewatripont (2002) and Laffont and Martimort (2002).

costs below the allowed level. In practice, as described in Chapter 2, the regulator shares a proportion of operating and capital efficiency rents with consumers, and thereby introduces a distortion between the theory of price cap regulation and the practice[17]. The sharing rule is decided without any explicit consideration of the firm's long-term reaction to the rule. We therefore expect that the rule is time inconsistent and that it will be ex-post inefficient[18].

We examine the impact of the sharing process on the firm's incentive to reduce costs. We assume first that the firm knows the sharing rule when making its effort choice. This allows us to focus on the impact of sharing itself. We then examine the additional impact of the firm's uncertainty about the sharing rule. Finally, the endogenous relationship between the chosen effort level and the sharing rule is considered. The focus in this section is on the detrimental impact of sharing for technical efficiency. We recognise, however, that sharing is advantageous for allocative efficiency[19].

We examine the impact of sharing by focusing on the firm's operating effort choice. Parallel results hold for the firm's capital effort decision. In our two-period model the price caps are calculated as[20]:

$$\begin{aligned} \widehat{p}^1 &= \theta - \widehat{e}^1 \\ \widehat{p}^2 &= \theta - \lambda_o^2 e^1 - \widehat{e}^2 \end{aligned}$$

The firm's profit level is:

$$\pi = \left(e^1 - \widehat{e}^1\right) q^1 - \psi\left(e^1\right) + \beta\left(\left(\left[1 - \lambda_o^2\right] e^1 + \left(e^2 - \widehat{e}^2\right)\right) q^2 - \psi\left(e^2\right)\right)$$

and the level of welfare is:

$$W = V(\widehat{p}^1) + \beta V(\widehat{p}^2) + \alpha\pi$$

[17]This is also recognised by Markou and Waddams Price (1999) who emphasise that 'As the process has developed some incentive power has been sacrificed as both regulators and companies have come to recognise that high revealed profits are likely to be reflected in a tighter price constraint at the next review'.

[18]A detailed analysis of time inconsistency problems can be found in Kydland and Prescott (1977). They argue that decisions which are selected on the grounds that they are best 'given the current situation and given that decisions will be similarly selected in the future' will be inefficient ex-post because no account is taken of the behavioural and expecation changes in response to the policy decision. The regulator's decision-making process is undertaken in a similar way and, hence, similar problems are to be expected.

[19]An analysis of the benefits of cost sharing can be found in Schmalansee (1989).

[20]We ignore the possible indirect effects of the firm's effort choice on the next period's price cap. For example, we do not allow the regulator's operating effort target for period 2 to depend on the actual effort level in period 1. This case, which arises in practice, would add further complexity to the comparisons being made.

where $V(p)$ is consumer surplus at price p, and α is the weight placed by the regulator on the firm's profits[21].

We compare the firm's effort choice, with and without sharing, to the welfare-maximising level. When there is no sharing the second period price cap is:

$$\widehat{p}^2 = \theta - \widehat{e}^1 - \widehat{e}^2$$

The first-order condition for welfare maximisation when there is not sharing is[22]:

$$\begin{aligned}\frac{\partial W}{\partial e^1} &= \alpha\left[q^1 - \psi'\left(e^1\right) + \beta q^2\right] = 0 \\ &\Rightarrow \psi'\left(e^1\right) = q^1 + \beta q^2\end{aligned}$$

The first-order condition for profit maximisation when there is no sharing is:

$$\begin{aligned}\frac{\partial \pi}{\partial e^1} &= q^1 - \psi'\left(e^1\right) + \beta q^2 = 0 \\ &\Rightarrow \psi'\left(e^1\right) = q^1 + \beta q^2\end{aligned}$$

We therefore confirm that, when there is no sharing, the firm chooses the welfare-maximising level of effort.

We now compare the effort levels when sharing is introduced. For a given sharing rule, λ_0^2, the welfare-maximising effort level is determined by the first-order condition:

$$\begin{aligned}\frac{\partial W}{\partial e^1} &= \beta\frac{\partial V(\widehat{p}^2)}{\partial e^1} + \alpha\left[q^1 - \psi'\left(e^1\right) + \beta\left[1 - \lambda_o^2\right] q^2\right] = 0 \\ &= \beta q^2 \lambda_o^2 + \alpha\left[q^1 - \psi'\left(e^1\right) + \beta\left[1 - \lambda_o^2\right] q^2\right] = 0 \\ &\Rightarrow \psi'\left(e^1\right) = q^1 + \beta q^2 + \left(\frac{1-\alpha}{\alpha}\right)\beta\lambda_o^2 q^2\end{aligned}$$

The welfare-maximising effort level is higher with sharing than without. The difference between the effort levels is increased the higher the level of sharing that occurs. This is not surprising as, for the given value of α, each unit of effort is worth more to the regulator the more sharing that occurs. This is because, with the weighted welfare function, the regulator places more value on the effort rent which is transferred to consumers than the rent which is retained by the firm. If $\alpha = 1$, so that the regulator weights profits and consumer surplus the same, the welfare-maximising effort level is the same with and without sharing.

[21] We assume that this weight is exogenous and that it is fixed in the current period. The value can change over time, however, if the regulator changes his perspective on the importance of the distribution objective.

[22] The second-order condition is $-\psi''(e)$ which ensures a local maximum with our convex cost of effort function. This holds for all the welfare and profit maximisations in this section.

The firm's profit-maximising effort choice with sharing is determined by the condition:

$$\begin{aligned} \frac{\partial \pi}{\partial e^1} &= q^1 - \psi'\left(e^1\right) + \beta\left(1 - \lambda_o^2\right) q^2 = 0 \\ &\Rightarrow \psi'\left(e^1\right) = q^1 + \beta q^2 - \beta \lambda_o^2 q^2 \end{aligned}$$

The higher the level of sharing that occurs, the lower is the level of effort chosen. This reflects the fact that each unit of effort is worth less to the firm the more sharing that occurs.

The firm's effort choice with sharing is lower than the effort choice with no sharing. It is also lower than the welfare-maximising choice *for the given sharing rule.* We therefore conclude that the existence of a sharing rule distorts the firm's effort choice away from the welfare-maximising level, and results in the firm undertaking a lower level of effort than is the case with no sharing. Technical efficiency is therefore reduced when savings are shared with consumers. This arises from the fact that the regulator has *multiple objectives.*

We note, as an aside, that the sharing of cost savings under the RPI-X mechanism has a similar impact to an explicit profit-sharing rule[23]. This is stressed by Armstrong and Sappington (2002) who argue that 'implicit intertemporal profit sharing of this sort can limit the firm's incentive to reduce its operating costs and expand its revenues, just as explicit profit-sharing requirements can'. The main difference between these regimes is that with formal profit-sharing the rule is set ex-ante, it is known to the firm, the proportion of rent which is shared is fixed over time, and sharing occurs immediately (ie, at the end of the financial year). In contrast, the implicit 'profit-sharing' under the RPI-X mechanism is built on an uncertain sharing rule, and the sharing occurs with a lag. The impact of this uncertainty, arising from the absence of commitment to the rule, is discussed below.

The impact of uncertainty The regulator chooses the sharing rule after the firm has made its effort choice, and makes no commitment from one review to the next as to what the sharing rule will be. The firm is therefore uncertain about what the rule will be when making its effort choice and its determines an expected sharing rule using the following function (see section 2.3.1 for details):

$$E\left(\lambda_o^{j,j+1}\right) = \lambda_o^{j-1,j} + \mu^j + v^j(e^j)$$

[23]See Burns, Turvey and Weyman-Jones (1995a and 1995b), Lyon (1996) and Waterson (1995) for a discussion of the properties of sliding-scale regulation. See Mayer and Vickers (1996) for an analysis of the problems with these schemes. Sappington (2000) presents details of the use of earnings-sharing schemes in the US telecoms sector.

Each period's sharing rule reflects the regulator's current position on the appropriate trade-off between distribution and technical efficiency. Several factors, unknown to the firm, affect this position. These include the impact of political and consumer pressure on the regulator; the size of the cost reductions delivered (introducing circularity into the firm's problem); and, recent decisions made by other regulators and the Commission on the appropriate level of sharing. The firm cannot predict the regulator's objective function and, hence, cannot determine the expected sharing rule accurately. There is thus a positive probability that the expected sharing rule will be incorrect.

We compare the effort choice with uncertainty to that which would emerge with certainty. To simplify the analysis we assume that the cost of effort function is $\psi(e) = \frac{e^2}{2}$. We therefore have:

$$\begin{aligned} e^j &= q^j + \beta q^{j+1} - \beta E(\lambda_o^{j,j+1}) q^{j+1} \text{ with uncertainty} \\ e^j &= q^j + \beta q^{j+1} - \beta \lambda_o^{j,j+1} q^{j+1} \text{ with certainty (ex-post)} \end{aligned}$$

The welfare-maximising effort level for the given sharing rule is: $e^j = q^j + \beta q^{j+1} + \left(\frac{1-\alpha}{\alpha}\right) \beta \lambda_o^{j+1} q^{j+1}$.

The firm's effort level deviates from the ex-post profit-maximising level if the expectation of the sharing rule is incorrect. Uncertainty about the sharing rule therefore results in ex-post inefficiency. In addition, the firm's effort choice will continue to be different from the welfare-maximising level. The extent of the difference depends on whether the firm has a pessimistic or an optimistic expectation of the sharing rule.

- *Pessimistic expectation*: if $E(\lambda_o^{j+1}) > \lambda_o^{j+1}$, the firm's effort choice will be lower with uncertainty than with certainty. This will widen the gap between the firm's effort choice and that which maximises expected welfare, given the actual sharing rule.

- *Optimistic expectation:* if $E(\lambda_o^{j+1}) < \lambda_o^{j+1}$, the firm's effort choice will be higher with uncertainty than with certainty. This will reduce the gap between the firm's effort choice and that which maximises expected welfare, given the actual sharing rule[24].

The firm is most likely to be pessimistic about the sharing rule if the regulator has increased the sharing rule from one review to the next. The firm adjusts the previous sharing rule upwards by the correction factor, μ_o^j, if it expects the regulator to tighten

[24]This raises the question of whether the regulator has an incentive to explicitly encourage the firm to have an optimistic expectation, perhaps by openly implying that a low sharing rule will be used at the next review. This additional level of strategic decision-making is not anaylsed here but may warrant consideration in future research.

the sharing regime over time. The larger the correction made, the more likely it is that the firm's expectation will overshoot the actual rule used by the regulator. In this way, the regulator's propensity to increase the rule over time increases the likelihood that the firm will have a pessimistic expectation.

The probability that the firm will have a pessimistic expectation is also increased if the regulator has shown, in previous periods, a propensity to revise the contract mid-period. The effect is expected to be cumulative, with one-off changes having only a small impact on the expected sharing rule, but ongoing and/or large changes having a larger impact. Again, therefore, the regulator's actions in earlier games impacts on the firm's expectations and hence on its current, and future, effort choices[25]. Finally, the propensity to be pessimistic will increase if the firm is risk averse.

We conclude that the firm's uncertainty about the sharing rule means that the effort choice will be ex-post inefficient. In addition, the effort choice will be significantly lower than the welfare-maximising level if the firm's expectation of the sharing rule is pessimistic. This is most likely to occur if the firm expects the rule to be reduced over time, either at future reviews or by increased mid-period intervention.

Additional ratchet effect In the above discussion we assumed that the regulator chooses the size of the sharing rule exogenously. In practice the situation is likely to be more complex, with the regulator adjusting the size of the sharing rule in reaction to the size of the cost savings made by the firm. When small savings are made, the need to maintain technical efficiency incentives is likely to be stronger than the need to distribute efficiency rents to consumers. When large cost savings are observed, however, distributional concerns are likely to become more prevalent. The level of effort therefore affects the way in which the regulator prioritises his objectives, and we can characterise the sharing rule as a function:

$$\lambda^{j+1}(e^j) \text{ such that } \frac{\partial \lambda^{j+1}(e^j)}{\partial e^j} > 0$$

In this situation the firm will also consider the impact of its effort choice on the future sharing rule. The net benefit of a high share of a smaller level of profit will be traded-off against the net benefit of a smaller share of a higher level of profit. There is thus a ratchet effect, with the firm's incentive to reduce costs, in order to earn rent, being offset by the risk that the rent will be reduced, via the sharing rule, if the level of cost reduction is too high.

[25]This point is also made by Laffont and Tirole (2000) who stress that 'early regulatory reviews further decrease the real power of formally high-powered incentive schemes'. Helm and Rajah (1994) also note, in relation to intervention by the water regulator, that 'The experience of frequent interventions in the first period will serve to weaken the credibility of the DG's fixed-period contract in the second'.

The probability of mid-period intervention also increases with the level of cost savings. The higher the cost savings, the more pressure the regulator will come under from the government and consumer groups to deliver price reductions as early as possible, and hence the higher the probability of intervention. Again, therefore, the firm considers the impact of its effort choice on mid-period intervention, and hence on the implicit expected sharing rule[26]. There is thus a secondary ratchet effect with this incentive mechanism.

We conclude that cost reduction incentives are dulled by uncertainty about the sharing rule. This occurs for the following reasons.

1. *Regulatory discretion* - the firm cannot predict the expected sharing rule accurately. This is because it cannot determine how the regulator sets the rule from one review to the next. There is therefore a positive probability that the expectation will be incorrect and the profit-maximising effort choice will be ex-post inefficient.

2. *Limited foresight* - the regulator does not consider the firm's reaction in future periods to current regulatory decisions. He therefore does not explicitly take account of the fact that a tightening of the sharing rule will increase the firm's expected sharing rule and, hence, reduce the level of effort chosen. In contrast, the firm considers the impact of current choices on the next period's sharing rule. The different horizons over which decisions are taken means that neither player accurately determines the other player's optimal actions and reactions, and decisions are not welfare-maximising.

3. *Repetition* - historic cost information is used by both the firm and the regulator. The sharing rule is changed in response to the level of savings of made. There is therefore a ratchet effect, where the firm reduces the effort level to minimise the probability of the sharing rule being increased. The firm also uses experiences at previous reviews as the basis for setting its expected sharing rule. The regulator's decision to increase the sharing rule over time therefore results in an increase in the firm's expectation, and hence a reduction in the effort level.

Use of historic information

In theory, a pure price cap is set without reference to the firm's current level of costs. Under the RPI-X mechanism, the regulator makes explicit use of current and historic cost information when setting the firm's price cap. This confirms Sibley's (1989)

[26]Vickers and Yarrow (1988a) note that the 'firm has to balance its desire for short-run profits against the risk of jeopardizing future profits by triggering a review of its prices'.

concern that regulators would be unable to commit to not using the data once it became available, even though it is optimal to ignore the information. The practice is therefore, once again, different from the assumptions underlying the theory, and technical efficiency incentives are dulled.

Historic and current information acts as a signal about future costs. The endogenous signal is of a relatively high quality, assuming some correlation between past and future costs, and hence the regulator is able to improve allocative efficiency by basing the price cap on a better information set. This improvement occurs at the expense of technical efficiency, however, and we see again the problems which arise when the regulator has multiple objectives. This trade-off is explicitly emphasised by Sappington (2000) who argues that using actual cost information to calculate the X-factor ‘can limit incentives for superior performance’ but it can also ‘limit the risk of affording enormous profit to the regulated firm or jeopardizing its financial integrity’.

We explain here how the regulator's use of current and historic information reduces technical efficiency. We know, from section 2.3.1, that the firm considers the impact of its effort choice on the next period's price cap. In the case of operating effort, the optimal effort level satisfies:

$$q^j + \beta(1 - E(\lambda_o^{j,j+1}))q^{j+1} - \psi'(e^j) + \beta\left(\frac{\partial E(\widehat{p}^{j+1})}{\partial e^j}\right) = 0$$

When assessing the size of the last term, $\frac{\partial E(\widehat{p}^{j+1})}{\partial e^j}$, the firm bases its expectation on the past decisions made by the regulator. In particular, if the regulator set forecast base operating costs equal to current operating costs at the last periodic review, the firm will assume that a similar methodology will be used at the next review. The firm therefore knows that by revealing information about its cost saving potential, it provides the regulator with the opportunity to reduce future rents. This will reduce the marginal net benefit from increasing effort, and will thereby dull the firm's cost saving incentives.

There is thus a classic ratchet effect problem in this dynamic game[27]. This effect arises because the firm makes decisions over a rolling two-period horizon. The problem would not exist if, like the regulator, the firm only considered the impact of its decisions on the current period[28]. This is recognised by Waddams Price (2000) who argues that incentives ‘proved to be effective within the price-cap period, but only for a myopic company which did not foresee the roles which lower costs or an inflated capital base

[27]The reader is referred to Weitzman (1980) for a ‘multiperiod stochastic’ analysis of the ratchet principle.

[28]Currie, Levine and Rickman (1999) suggest that the effect could also be reduced if a pro-industry regulator was appointed by the government. This would have significant implications for welfare and distribution, however, which may outweigh the gains from reducing the ratchet effect.

might play in determining the next cap'.

The ratchet effect is reduced if the regulator uses information on average costs in the sector to assess the firm's future costs. The firm's price is then only partially dependent on its own costs, and the impact of current effort decisions on the next period's price cap is reduced. This is consistent with the idea that mechanisms based on yardstick competition improve efficiency incentives. The ratchet effect is not completely removed, however, because the comparator information is generally used alongside the firm-specific information, rather than in its place.

We conclude that the cost reduction incentives under the RPI-X mechanism are reduced because current and historic cost information is used to the set price cap. The problem exists because of the following three characteristics of the RPI-X game.

1. *Multiple objectives* - the desire to improve allocative efficiency, by using a better information set, is traded-off against the reduction in cost-minimisation incentives.

2. *Limited foresight* - the firm takes account of its decisions on the next period's price cap. The regulator, with his one-period decision horizon, does not consider these on-going effects when making the trade-off between allocative and technical efficiency.

3. *Cost observability* - historic cost information is easily available. The regulator does not, however, have reliable information on future costs or on the firm's underlying cost functions. The historic data is therefore the best information set available.

There are two ways in which this problem might be alleviated. First, as noted by Armstrong and Sappington (2002), 'When regulators cannot make binding commitments regarding their use of pertinent information, welfare may be higher when regulators are denied access to the information'. It may therefore be better to make it harder for the regulator to access the information, perhaps by requiring the regulator to pay a fee to the firm, or an external body, for the data. Alternatively, the data could be made available with a lag, so that current information could not be used by the regulator. Both these options are problematic, however, as they remove the allocative efficiency gains of basing the price cap decision on the best available information set. They also run contrary to current legislation which requires private corporations to publish annual accounting reports. An alternative potential solution to the problem is, as briefly noted above, to base the firm's price cap on the current costs of other firms in the sector. Such relative schemes are considered in more detail in section 3.2.1.

Input choices

We have shown that, under the RPI-X mechanism, the firm has an incentive to reduce costs. This incentive is dulled by the use of an uncertain sharing rule and by basing the price cap on historic and current cost information. However, despite these limitations, some cost savings are expected to be delivered under the mechanism. We consider here the possibility that the reduction in costs will not necessarily improve welfare. This occurs if a decrease in one type of costs is simply offset by an increase in another type of cost (eg, labour costs are replaced with capital maintenance costs), or if the decreased costs are associated with a reduction in output produced. In this sense, cost savings are not always consistent with technical efficiency gains.

Operating and capital investments are often substitutable in the utility sectors. The firm may be able to reduce one type of investment by increasing the other[29]. These choices can be made efficiently by the firm, in a standard cost minimisation exercise. The RPI-X mechanism may distort the firm's input choices away from this efficient level however[30]. This arises because the incentive mechanisms used to induce increased effort differ, in their impact, between operating and capital costs. The relative value of capital to non-capital inputs is therefore changed by the regulatory regime.

The two main reasons for this distortion are summarised here.

1. Assume that $\widehat{e}^j = \widehat{\gamma}^j = 0$ and $e^j = \gamma^j = 1$. The efficiency rents from operating and capital effort are then:

$$\begin{aligned}\Delta\pi^o &= q^j + \beta(1 - \lambda_o^{j,j+1})q^{j+1} - \psi(1) \\ \Delta\pi^c &= r^j + \beta r^{j+1}\lambda_c^{j,j+1} - \psi(1)\end{aligned}$$

   Given that the interest rate is less than one, and the sharing rules are chosen independently, we expect the value of a unit of operating effort to be higher than the value of a unit of capital effort[31]. This reflects the fact that the firm only earns the financing cost of capital savings under the annual price cap but it

[29] The evidence on productivity improvement in the utility sectors since privatisation (see Chapter 4) suggests that such input substitution has occurred. In most industries labour productivity has increased while total factor productivity has declined, or increased at a slower rate. This suggests, according to Saal and Parker (2001), that 'Capital for labor substitution has been occurring during the 1990s'.

[30] Frontier Economics (2003b) provides a discussion of the input distortions which arise with the mechanisms used by both Ofwat and Ofgem.

[31] Frontier Economics (2003b) find that 'Direct comparison of the benefit to companies from a one-off (recurring) gain in operating expenditure against the benefit of a one off (recurring) gain in the capital expenditure suggests that companies face stronger incentives to make operating expenditure savings'. They argue that 'economic efficiency is served....through setting equal incentives and allowing companies to pursue the least total cost option'.

earns the total level of operating cost savings. The firm is therefore more likely to make operating cost savings than capital cost savings, and its input decisions are distorted away from the cost-minimising choices of an unregulated firm.

2. Capital investment is retained in the regulatory capital value over time, and the regulator is required to allow the firm to earn an adequate return on this investment. There is thus an incentive to overcapitalise when the allowed cost of capital is higher than the firm's actual cost of capital[32]. High capital levels can be induced by overinvesting in the assets, or by not investing in capital effort. In either case, the firm's incentive to reduce capital costs is dulled and the price-setting process may result in input distortions similar to those associated with rate of return regulation (ie, the Averch-Johnson (1962) effect). This incentive to 'raise the capital base to increase allowed profits' is also noted by Markou and Waddams Price (1999).

The RPI-X mechanism may also distort the timing of input investment decisions[33]. When savings are shared at the periodic review, the firm has an incentive to make cost savings early in the period to reap the maximum benefit from those savings. In addition, because current cost information is used to set base operating costs for the next period, the firm has an incentive to have higher costs at the end of the period. This means that input decisions are influenced by the timing of the next review and are not solely driven by cost minimisation considerations over time. There is evidence that such trade-offs have been made by firms, with large reductions in operating costs just after a price review and increased costs as the next review approaches.

This discussion suggests that the way in which the RPI-X mechanism is designed distorts the firm's input choices. The welfare cost arising from the inefficient use of inputs may offset the technical efficiency gains arising from cost reductions. The problem does not arise because of any fixed characteristics of the RPI-X game and could be changed by altering the details of the mechanism. For example, both the water and electricity regulators attempted to deal with the timing distortion problem in the 1999 price reviews. Rolling sharing rolls, where the length of the regulatory lag is independent of the time that savings are made, were introduced[34]. These rules are discussed in Appendix A. More generally, regulators could consider the impact of

[32]The overall effect depends on whether the profit gain from high capital levels, arising from the difference between the allowed and actual cost of capital, is higher or lower than the profit gain from investing in capital effort. The trade-off will depend on the length of the regulatory lag and the extent to which the allowed return is higher than the cost of financing the investment.

[33]See Pint (1992) for an analysis of the impact of the regulatory regime on the timing of investment.

[34]Ofwat, the water regulator, introduced a rolling rule for operating and capital costs. Ofgem, the electricity distribution regulator, only introduced the rule for capital costs. This may have increased the distortion between operating and capital cost incentives further.

the regulatory mechanism on total costs, and input choices, rather than focusing on reductions in operating costs and capital costs separately[35].

3.1.3 Efficient output delivery

The level of welfare is affected by the amount of output that the firm produces, and by the quality of the output which is delivered to consumers. We consider here whether the firm has appropriate incentives, under the RPI-X mechanism, to undertake the long-run investment required to ensure that sufficient capacity is always available. We also examine the firm's incentive to deliver efficient levels of quality of service. The main concern is that the firm has a number of tasks to deliver - cost reduction and output delivery - and the incentive schemes which are focused on each individual task may conflict.

Investment incentives

Under a pure price cap regime investment costs are completely separate from the allowed price, and the regulator makes no commitment that long-term costs will be recouped through the price cap. The more concerned the regulator is about distributing benefits to consumers today, the higher is the risk that the allowed return will be lower than the required return. In this setup the firm will be reluctant to undertake any long-term investments as it will be concerned that the regulator will not allow it to earn an adequate cost of capital. A welfare-maximising regulator will indeed have an incentive to 'hold-up' the firm in this way once the investment has been sunk[36]. This is confirmed by Vickers and Yarrow (1988a) who argue that 'investment is inhibited by the fear of 'unfair' future regulation. Unless the regulatory and/or political systems provide a credible means of commitment of future fairness, this fear is not entirely unreasonable in view of the objectives of regulators'.

In the RPI-X game the regulator does not adhere to the strict principles of pure price cap regulation. In particular, the regulator provides the firm with some commitment on the return which it will be able to earn on long-term investments. The

[35] It would be interesting to focus, in future research, on the regulation of total investment with a multi-dimensional effort variable. This would capture the operating and capital elements of total investment, and may allow for the design of incentive mechanisms which take account of the trade-off across these categories.

[36] Armstrong, Cowan and Vickers (1994, section 3.6) provide a useful overview of this hold-up problem. Armstrong and Sappington (2002) emphasise that the regulator's ability to 'hold-up' the firm can be limited by 'imposing a legal requirement that the firm make a specified return on its assets'. A theoretical model of the firm's optimal investment decision in these circumstances is provided by Besanko and Spulber (1992). The hold-up problem under pure price cap regulation is also discussed in Gilbert and Newbery (1994).

hold-up problem is partially alleviated in the following ways[37].

1. The regulator has a legal duty to ensure that the firm always earns a reasonable return on its asset base. Where the firm deems that the regulator has not set an adequate rate of return, it has the option of seeking an appeal to the Commission[38].

2. The price cap is set explicitly to ensure that allowed revenues in the next period cover expected costs. If the firm considers a long-term project as a series of 'short-term' phases, it can attempt to ensure that the financing of each phase is adequately recovered at each periodic review[39]. In this regard Armstrong, Cowan and Vickers (1994) argue that 'Commitment to future prices, if feasible, would be desirable ex ante. However, if the firm and regulator interact repeatedly, relatively efficient investment decisions can be encouraged'. A series of short-term contracts may also be preferable if exogenous shocks, and stochastic demand, mean that long-run investment requirements are highly uncertain.

3. Risky projects, where the costs fluctuate significantly because of exogenous factors, may be avoided by the firm. To overcome this problem, regulators have used cost pass-through schemes for costs which are considered to be out of the firm's control. This ensures that investments which are uncertain at the time of the regulatory review will be automatically included in the capital value once incurred.

The RPI-X regime therefore incorporates elements of rate of return regulation which reduce the firm's concern about an investment hold-up problem. Inevitably, as some of these provisions increase the link between price and actual costs, there will be an off-setting reduction in technical efficiency incentives. In addition, the regulator will only allow investment in the regulatory capital value which is considered necessary and efficient. The firm will therefore remain reluctant to undertake discretionary

[37] Levine (1999) presents a more detailed discussion of the type of solutions which are proposed in the literature to overcome the hold-up problem. The author suggests that the problem can be alleviated if the government appoints a regulator who is known to be 'pro-industry'. This would, of course, lead to its own problems with respect to decisions on allocative efficiency, technical efficiency and distribution.

[38] Besanko and Spulber (1992) propose that an extreme rate of return guarantee should be provided 'that is not revised by the regulator after the firm has made its capital investment'. The current 'financing of functions' duty does not provide such a long-run guarantee and, hence, may not induce optimal long-run investment levels.

[39] The outcome is similar to that which arises in Salant and Woroch's (1992) optimal pricing policy. In this model, the regulator agrees to a specific time path for capacity investment and agrees to cover the costs of each phase as they are being undertaken (see Armstrong, Cowan and Vickers (1994, pp88-89)).

investments which are not directly associated with current output requirements[40]. We also note, as found in the gas sector, that the regulator may allow an adequate return on the asset base, but may not guarantee a depreciation schedule which ensures that maintenance investment is adequately financed[41].

We conclude that, in contrast to our discussion on technical efficiency, the specific details of the RPI-X mechanism may increase welfare relative to the theory of pure price cap regulation. The risk of the hold-up problem is reduced, although it is not completely removed. Despite this, underinvestment may still arise. This is because of the cost-saving incentive which is paramount in the RPI-X mechanism. Total costs can be reduced by simply not undertaking the required work. In this way the incentive to improve technical efficiency works against the objective of ensuring that efficient output levels are delivered[42]. This outcome is synonymous with the predictions of the literature on multi-task incentive schemes[43]. We discuss this problem with respect to the delivery of quality of service below.

Meeting quality of service targets

A 'command and control' approach is used to regulate the quality of service delivered to consumers[44]. A target is set and the firm is provided with a financial incentive

[40]To offset this problem with long-term projects regulators in the water and airports sectors have allowed the firm to include investment on assets in the course of construction in the asset base, rather than requiring the firm to demonstrate that outputs have been produced before financing the investment.

[41]During the 1997 review of Transco's price cap Ofgas proposed using a 'pay-as-you-go' depreciation schedule, where depreciation charges would be increased at the time when the investment is required. The firm, in contrast, sought a phased depreciation schedule, where the costs of investing in the asset base were recovered over the life of the asset. As noted by Waddams Price (1997), concerns were raised about whether, under the 'pay-as-you-go' scheme, the regulator would be able to commit future regulators 'to allowing such revenues over several future price reviews'. Concerns about the regulator 'holding-up' the firm once the asset base has been sunk are therefore evident in practice.

[42]There is some evidence that the profit-incentive has led to a reduction in investment in output delivery. For example, Markou and Waddams Price (1999) argue that investment patterns for British Gas indicate that 'short-term profit improvement' may motivate the company to lower its investment levels. They also argue that, in the water sector, 'a decline in renewals and maintenance of the existing distribution network have proved an important source of dividend growth'.

[43]Bennett and Waddams Price (2002) also discuss the problems which arise with multiple objectives in the regulatory field. In a more general discussion, Helm and Yarrow (1988) emphasise that 'constraints on any one of the regulated firm's decision variables is likely to lead to adjustments in other decisions'. Laffont and Tirole (1993, Ch3) consider the multi-task problem in the more general framework of designing a contract for regulating a multiproduct firm, where quality is one of the products produced. Theoretical analysis of multi-tasking can be found in Athey and Roberts (2001) and Holmström and Milgrom (1991).

[44]We focus here on the regulation of quality of service in the water and electricity distribution sectors. In these sectors quality of service regulation is considered as part of the RPI-X regime. In other sectors additional regulatory mechanisms exist for quality of service which operate outside the price cap regime. For example, sliding-scale incentive schemes exist for the system operator business of the National Grid Company. Regulators may benefit from exploring these schemes further to consider

to meet that target. We assume here that the target set is appropriate and consider whether the firm has an incentive to meet the target in the RPI-X game. In the next section we consider whether the target is in fact efficient.

If there is price cap regulation, and no associated output regulation, the firm will have an incentive to reduce the quality of service delivered. This is because a reduction in quality of service is synonymous with a reduction in costs and the firm is able to earn above normal profits, at least until the end of the regulatory period. This incentive may be offset, to some extent, by the incentive to increase capital investment in the asset base and thereby earn higher profits in the long-run[45]. The net effect depends on the difference between the profit gained from a higher capital value, and the profit gained from reducing costs below the allowed level.

An analysis of BT's service performance after privatisation suggests that the incentive to reduce costs was stronger than the overcapitalisation incentive[46]. Laffont and Tirole (1993) also note that in the US 'there has been concern that 'incentive regulation'.....conflicts with the safe operation of nuclear power plants by forcing management to hurry work, take short-cuts, and delay safety investments'. The problem of non-delivery of outputs arises, according to Williamson and Mumssen (2000) 'because regulation seeks to constrain prices, rather than as a consequence of monopoly per se'. Laffont and Tirole (1993) also argue that the higher the incentive to make cost savings, the greater is the incentive to reduce the level of outputs delivered. This outcome is consistent with the conclusions reached in Holmstrom and Milgrom's (1991) analysis of multi-task principal-agent models.

Regulators have attempted to address this problem by introducing output regulation alongside the price cap mechanism[47]. The question we wish to consider is whether the schemes provide sufficiently strong incentives for the delivery of an efficient level of quality of service[48]. We first consider the impact of the quality of service regimes which existed between privatisation and 1999. In this timeframe, the regulatory tools used were consumer compensation payments and the 'name and shame' approach.

whether they present viable, and welfare-improving, alternatives to the regime described here.

[45]This overcapitalisation incentive is emphasised in Waddams Price, Brigham and Fitzgerald (2002).

[46]See Armstrong and Vickers (1995) and Rovizzi and Thompson (1995) for a description of the service problems which arose. Markou and Waddams Price (1999) also provide a review of the concerns which arose about quality of service when price cap regulation was introduced for British Telecom and British Gas in the 1980s. Banerjee and Dasgupta (2001) provide a more recent analysis of the relationship between quality of service and incentive regulation in the US telecoms sector.

[47]We note that the solution used is to introduce separate incentive schemes for cost reduction and output delivery, rather than to design mechanisms which focus on both tasks simultaneously. This conflicts with Holmström and Milgrom's (1991) argument that 'incentive problems must be analyzed in totality'. Further research on the application of optimal multi-task incentive mechanisms to the regulation field may provide useful ideas on how to improve incentives in both areas.

[48]Rovizzi and Thompson (1995) provide a complementary analysis of the advantages and disadvantages of the various schemes which are used to regulate quality of service.

These tools are expected to have a weak impact on output delivery for the following reasons.

1. Under the compensation schemes the level of payment per customer is low and the number of customers who make claims is also likely to be low[49,50]. Firms therefore do not face a significant financial cost for choosing a lower level of service - ie, $C\left(s_l^j\right)$ is low.

2. The 'name and shame' approach changes the relative size of the reputation parameters in the firm's decision function. It is not clear that the firm would place a high weight on this impact relative to the profit made from underinvestment. This is particularly true when we consider that these are monopoly firms who do not face the risk of losing customers, even if they have a poor reputation for quality of service.

3. The regulator is not always able to impose compensation payments in practice. This is because outputs are not always observable. Furthermore, low investment levels can not be associated directly with low output levels. This is because other factors, particularly cost efficiency improvements and exogenous shocks, affect the relationship between output delivery and required investment levels. It is therefore difficult to implement the quality of service regulatory regimes in practice.

Despite the theoretical concerns with these regulatory mechanisms there has been no evidence of significant service problems in the utility sectors during this period[51]. For example, according to Waddams Price, Brigham and Fitzgerald (2002), 'The performance of the water and sewerage companies against the various DG measures shows a pattern of consistent improvement across all companies and measures'. Similarly, in the electricity distribution sector, 'There are steady and consistent improvements'.

[49]Frontier Economics (2003a) found that payments in the electricity distribution sector 'begin at £50 (for domestic customers) 18 hours after supply fails, with further payments of £25 for every subsequent 12 hours until restoration. In a typical year the sums involved are not large in comparison to total price control revenue - tens to hundreds of thousands of pounds (price control revenue is typically £100-200 million per year)'.

[50]Markou and Waddams Price (1999) note that, in the early years after privatisation, few customers in the water sector made claims under the guaranteed standards compensation schemes. Indeed, the regulator believed that those who made claims represented 'only 1-2% of all the justifiable claims, and attributed the difference to customer ignorance (only 15% know of the scheme) and/or a belief that the compensation involved (£5 for most claims) was not worth it'. The proportion of customers seeking compensation has no doubt increased since then, but similar problems remain.

[51]Network Rail (formerly Railtrack) is the clear exception. In this case it seems that the elements of the RPI-X mechanism which provide the firm with an incentive to reduce capital costs outweighed the output delivery incentives. Further research is required to determine why Railtrack's output delivery performance differed from that of electricity and water firms.

The quality of service record in the utility sectors may suggest that the reputation effect is stronger than expected and/or that the firm is motivated by other, non-profit, concerns. Waddams Price, Brigham and Fitzgerald (2002), in discussions with company managers, found that 'Companies are clearly sensitive to the standards set, but the direct financial penalty is often less important than reputational effects'. The reputation effect is particularly important for regulated firms expanding into international markets where performance is an important factor in determining whether or not they are awarded contracts[52]. The National Audit Office (2002) also found, in a survey of water and electricity company managers, that 'the publication of league tables by the regulator acted as an incentive for companies to maintain and improve service performance in order to maintain their reputation'.

There are concerns, however, that the current quality of service performance may not be an accurate reflection of future performance. This is because, as emphasised by The National Audit Office (2002), 'pipe and wire networks have an underlying resilience and it could take some time for inadequate or inefficient expenditure on maintaining them to be reflected in declining performance against output measures'. If this is the case, then weak incentives in the regime may result in service problems in the future. Finally, while no problems have emerged it is not clear that the quality of service has been as high as it might have been, or, indeed, whether it has been too high.

In the 1999 water and electricity distribution reviews, the regulators tightened the output regulatory regime by linking quality of service with the price cap. This change is expected to increase the firm's incentive to deliver a high quality of service, simply by making low quality of service more costly. A number of concerns remain, however, as to whether the adjustment offers maximum incentives to the firm.

1. The size of the adjustment is arbitrary and may be quite small.

2. The regulators can change both the basis on which the firm's performance is assessed, and the size of the adjustment, at future periodic reviews. The firm therefore has ongoing uncertainty about the implications of its decisions.

3. The adjustment is based on the assumption that service problems will be observed during the regulatory period and can then be penalised at the next periodic review. The lag between underinvestment and service problems may be

[52]Waddams Price, Brigham and Fitzgerald (2002) suggest that this reputation effect may be so strong as to result in the company producing too much quality. For example, a firm may wish to improve quality of service significantly to gain a reputation as an international utility company which provides quality. The firm will be particularly keen to do this if the costs of the investment can be recouped through the price cap. In this way, the authors argue, 'consumers in the regulated market are effectively paying for the companies' marketing activity elsewhere'.

longer than this, however, and may not be reflected in a price cap adjustment for some time. The higher the firm's discount rate, and the longer the expected lag, the less weight will be placed on the penalty by the firm.

4. The adjustment is based on the firm's performance relative to the target set by the regulator. It is not clear that this is the optimal incentive to provide, as the target may not represent an efficient level of service. This is discussed in more detail below.

5. Quality of service performance is affected by exogenous stochastic factors. For example extreme storms affect electricity availability on the transmission and distribution networks and hot dry summers limit the availability of water. To ensure that firms are not penalised for factors which are outside their control the price cap adjustment must distinguish firm-induced quality of service problems from those which are considered uncontrollable. This is important both for assessing performance relative to the firm-specific target and for assessing performance relative to rest of the sector when the exogenous shocks affect firms in different ways. This increases the complexity of the adjustment mechanism, which is highly aggregated in nature, and may increase the risks faced by the firm if not appropriately accounted for in the regulatory regime. Such risks would need to be compensated elsewhere - in higher required costs and/or a higher cost of capital - and may result in higher overall price caps.

As the price cap adjustment has only recently come into force, it is too early to assess whether the level of quality will improve under this tighter regime. Further research needs to be undertaken to assess the precise impact of the mechanism, and to determine whether particular aspects of its design affect the degree to which it improves quality of service.

Efficient targets

Rovizzi and Thompson (1995) argue that 'efficient resource allocation requires that standards be established with reference to consumers' valuation of a quality improvement and the corresponding costs of achieving it'[53]. This reflects the idea that a target is only efficient if it is equal to the level at which the marginal social benefit of the output equals the marginal social cost. We therefore consider an output target to be efficient if the associated price reflects the consumers' willingness-to-pay for the output level.

[53] Waddams Price, Brigham and Fitzgerald (2002) propose a similar approach to setting an efficient target: 'Economic theory suggests that regulators should choose standards according to consumers' valuation and the marginal cost of quality improvements'.

Given this definition of an efficient output target, we argue that the quality of service targets set by the regulator may not be efficient[54]. The regulator considers the trade-off between increased prices and increased quality of service, but he does not have explicit information on consumer willingness-to-pay, particularly at firm level. In addition, as emphasised by Markou and Waddams Price (1999), the regulator's decision on the appropriate target is based on vague duties. The standards therefore reflect, at best, the regulators' judgement on what the price-output trade-off should be. This judgement may, as noted by Waddams Price, Brigham and Fitzgerald, be 'motivated by political as well as economic factors'.

Even with a perfect incentive scheme, therefore, the firm may not deliver efficient output levels in the RPI-X game because the targets, which the firm is aiming for, may not be equal to the efficient level. This is a direct consequence of the limited information which is available to the regulator. It may be possible to obtain more efficient targets by improving the information set or, as discussed in section 3.2.2, by delegating the decision on the price-output trade-off to the player with better information on consumer willingness-to-pay. In this context we note that the firm may have better information than the regulator about this trade-off. More importantly, local consumer groups are expected to have the best information about average willingness-to-pay for particular outputs[55]. For all players, however, uncertainty about the required efficient level of quality will remain because of the stochastic nature of demand. Inefficiency will therefore remain, even if the target is chosen by the player with the best *current* information on average preferences.

3.2 Ideas on improving efficiency

Welfare is not maximised under the RPI-X mechanism. Allocative efficiency is not delivered, incentives for technical efficiency are dulled by the particular features of the RPI-X game, and, even with output regulation, it is not clear that the regulated firm is providing an efficient level of output to consumers. The mains reasons for these problems are outlined in section 3.1.

We therefore wish to determine ways in which welfare *might* be increased in the regulated sectors. We assume in this chapter that it is best to retain the RPI-X mecha-

[54]A similar point is made by Forsyth (1999) with respect to price cap regimes in Australia. He notes that 'The quality monitoring approach supposes that the initial level of quality is optimal; this may or may not be the case'.

[55]It is clear that customers are not homogeneous and hence will have different preferences about the appropriate trade-off between price and quality. It would be cumbersome and costly, however, to collect and collate information on the willingess-to-pay for every individual consumer. We therefore assume that it is only feasible to base any mechanism on an assessment of what the average customer wants. While not ideal this is the only realistic option in the absence of a competitive market.

nism, and we focus on ways in which its design and implementation can be changed[56]. This approach is justified on the grounds that it would be costly, particularly in terms of regulatory uncertainty, to introduce an entirely new mechanism. It is also not clear that a feasible alternative can be easily designed. Finally, technical efficiency gains have been delivered under the RPI-X mechanism, and it is important that these existing benefits are retained. Helm (1995c) also argues that 'Reforms are likely to be more successful if they build on the existing system, than if they require tearing it up and starting from scratch'.

Improvements in welfare can most easily be delivered in the short-run by focusing on the design of the price cap itself. In the long-run, it may be desirable to also change the characteristics of the game, but a cost-benefit analysis should be undertaken before proceeding in this direction. The elements of the regime which we consider changing are as follows.

1. Technical efficiency incentives are dulled by the way in which savings are shared from one review to the next. We examine the possibility of improving these incentives by changing the way in which sharing is undertaken.

2. The quality of service delivered is not necessarily efficient. We propose a way in which the regulator may get closer to the efficient level of quality of service. Once the standard is identified, the firm can then be provided with an incentive to meet the target.

We explain how our proposed changes to the RPI-X mechanism *might* improve allocative efficiency, technical efficiency and/or the delivery of an efficient level of output. We also, importantly, discuss the feasibility of the proposals given the characteristics of the RPI-X game and the uncertain nature of the firm's operating environment. No doubt several other improvements could be considered which relate to other facets of the regulatory mechanism[57]. These are left for further research.

Many of the proposed improvements have been suggested by others, usually in a more general context. The relevant literature is discussed alongside the individual proposals. Our contribution is to link the ideas with the specific features of the game described in Chapter 2, and to explain how the changes might yield a higher level of

[56]The need for change is supported by Mayer and Vickers (1986) who argue that 'There are good reasons for seeking reform of current price regulation: allocative efficiency, distributional fairness and the credibility of regulation can all be enhanced through changes to existing regulatory rules'.

[57]There are alternative ways of improving welfare in the regulatory regime which do not involve changes in the mechanism itself. Most notably, welfare could be increased by franchising the right to run the monopoly business (ie, competition 'for the market'). This option is discussed by Crampes and Estache (1997), Demsetz (1968) and Harstad and Crew (1999).

welfare than the existing RPI-X mechanism. In essence, we provide a new twist on existing ideas by emphasising how they apply to the current practice of regulation.

We conclude this introduction by emphasising that this is the *starting point* for the design of an amended RPI-X mechanism. The ideas are preliminary and high-level at this stage. In addition, each proposed change is considered in isolation. Future research will need to include an analysis of the impact of the proposals on each other, and on existing elements of the RPI-X mechanism. Regulators will also need to consider how to convert the proposals into feasible, perhaps sector-specific, amendments to the current price cap regime. A more detailed analysis of the information requirements of each proposal, and the implications of uncertainty arising from stochastic demand and cost conditions, will also be required.

3.2.1 Improving the sharing rules

Cost saving incentives are dulled in the RPI-X game because savings are shared with consumers at each periodic review. We accept here that some element of sharing needs to be retained in the RPI-X mechanism. This is because sharing ensures that allocative efficiency is at least partially restored at fixed intervals and, importantly, it ensures that the price cap regime is politically acceptable and hence sustainable in the long-term. Mayer and Vickers (1996) similarly find, in the context of automatic profit-sharing rules and a static model, that 'some relaxation of incentives is desirable because the loss of technical efficiency is second-order whereas the gain in allocative efficiency (or distributional efficiency, or risk sharing in a richer model) is first-order'. Sappington (2000) also argues that some sharing may increase aggregate welfare because the 'losses from diminished incentives to minimize production costs are small relative to the gains from better aligning prices and costs when small amounts of earnings sharing are introduced (Lyon, 1996)'. The net impact depends on the weight which the regulator places on the firm's profits relative to consumer surplus.

The way in which the sharing rules are designed has important implications for the firm's cost saving incentives. We found, in section 3.1.2, that the technical efficiency problems are exacerbated by the fact that the sharing rule changes over time. This is because the firm's effort choices are affected by uncertainty about the rule. We therefore suggest that technical efficiency incentives can be improved by requiring the regulator to commit to a fixed sharing rule. This rule should be chosen to maximise constrained welfare, given the existence of a sharing rule. Information limitations may prevent the regulator from finding this optimal fixed sharing rule however. We therefore also consider the familiar remedy of using a yardstick rule for setting the firm's allowed efficiency rent.

The welfare properties of the fixed sharing rule and the yardstick rule will be affected by the stochastic nature of the firm's operating environment. Even if the regulator had full information about current cost and demand conditions, it is unlikely that the ex-ante welfare-maximising rules would be ex-post efficient. This is because uncertain exogenous factors - notably weather - have a significant impact on demand in the water and electricity sectors[58]. By fixing the sharing rule the regulator loses the flexibility to react to such exogenous shocks and there may be a loss in allocative efficiency. In addition, where firms are affected differently by demand shocks the yardstick mechanism may not provide an appropriate benchmark. This is because the efficient costs of one firm may no longer by correlated with those of another firm. Uncertainty and the nature of the shocks to the firm's operating environment may therefore reduce the welfare gains which arise from the improvement in technical efficiency delivered by these alternative sharing rules.

Proposal 1: Fixing the sharing rule

Our first proposal is that the regulator should fix the size of the sharing rule over time - ie, set $\lambda_o^{j-1,j} = \lambda_o$ and $\lambda_c^{j-1,j} = \lambda_c \; \forall j$[59]. We found, in section 3.1.2, that when the firm is uncertain about the sharing rule, its effort choice will be ex-post inefficient. The firm is unable to make accurate forecasts of the rule because it cannot predict how the regulator's judgement will change over time. A fixed rule provides certainty and ensures that the firm's effort choice is efficient, given the value of the sharing rule[60]. We assume here that the rule is independent of the level of effort undertaken by the firm[61]. This further improves technical efficiency by reducing the ratchet effect. The

[58]The impact of the stochastic shocks vary somewhat by sector. In electricity, a severe storm may result in service unavailability for a relatively short period of time. The firm would incur increased costs when fixing the problem and reduced quality of service performance. In the water sector hot dry summers are the prime source of concern. In the event of a drought the water network is placed under severe pressure because of increased and peaky demand. This will result in service problems - including potentially prolonged periods of water shortages - and will lead to uncertain investment levels as companies attempt to alleviate the problem. In addition, water companies invest in increased capacity to minimise the risk of not being able to provide supply in the face of increased demand, particularly at peak times of the day. The weather-effect is therefore, to some extent, more extreme and more long-term in nature than that in the electricity sector.

[59]The idea of using a fixed sharing *rule* is consistent with Kydland and Prescott's (1977) conclusion that, when discretionary policy-making leads to time inconsistency problems, a rules-based approach should be used. The proposal is also similar to the 'Passive Target Setting' rules discussed by Laffont and Tirole (1993, Ch 9). With these rules dynamic contracts are designed to change in accordance with a known, and exogenous, revision rule over time.

[60]The effort level is not optimal - ie, welfare maximising - as the existence of any sharing rule, whether fixed or not, reduces the cost saving incentive.

[61]It may be possible to design a fixed sharing *function*, which allows the sharing rule to change, in a known way, in response to to the level of effort. This would provide the regulator with extra flexibility in terms of improving distribution and allocative efficiency, but would mean that the ratchet effect problems associated with sharing would remain.

proposal will only work, however, if the regulator can credibly commit to the fixed rule over time.

An exogenous mechanism must be introduced in the RPI-X regime to ensure that the regulator is committed to the fixed sharing rule[62]. One option is to explicitly incorporate the requirement of a fixed rule into the firm's licence and, perhaps, primary legislation[63]. The actual value of the rule should also be included in the licence. This ensures that all parties know the rule when making decisions, and it also means that the regulator requires the firm's agreement to change the rule in the future. If the principle of a fixed sharing rule is included in primary legislation it would require the consent of parliament to be changed.

We stress that what we require here is commitment to the fixed sharing rule only; not long-term commitment to the entire regulatory contract. The regulator retains the discretion to change the other elements of the price cap calculation at each periodic review[64]. This is important because the regulator can reap allocative efficiency gains by basing each new price cap on the regulator's improved information set. The adjustment to the cap will reflect changes in the underlying cost and output assumptions for the next regulatory period, rather than changes in the way in which historical savings are shared with consumers. It therefore more closely reflects the idea of a forward-looking price cap. Commitment to a sharing rule does, however, imply that there can be no mid-period intervention to share savings early with customers.

Our idea of 'partial commitment' reflects a compromise between two key results in repeated game theory[65].

1. With complete contracts, full commitment is optimal in a repeated game. In this context the 'the optimal long-term contract is obtained in a straight-forward manner as the replica of the one-shot optimal contract' (see Laffont and Martimort (2002)). Armstrong and Sappington (2002) stress that all other policies are sub-optimal since 'the regulator with full commitment powers could implement

[62]Crew and Kleindorfer (1996) suggest that the existence of known sharing rules, as are observed in several incentive schemes in the US, may be 'less vulnerable to reneging by the regulator'. This is because there is a transparent device by which customers always gain when the firm does well and which limits how well the firm can do. There may therefore be less political pressure on the regulator to implement changes. In this sense the fixing of the rule may act as its own commitment device.

[63]Armstrong (2003) argues that 'contracts or licences are the obvious way to prevent opportunistic behaviour'.

[64]Armstrong, Cowan and Vickers (1994) emphasise that regulatory reviews are beneficial as they allow the regulator to take account of the fact that 'Industry conditions - involving costs, new products, and the extent of competition - are likely to evolve in ways that are not foreseen'. There is no mention in this list, however, of the need to hold reviews to reflect changes in the regulator's methodology.

[65]The reader is directed to Armstrong and Sappington (2002), Laffont and Martimort (2002) and Laffont and Tirole (1993), and the references therein, for a general analysis of optimal regulatory contracts under a range of commitment options in a dynamic environment.

this policy, but chooses not to do so'[66].

2. With incomplete contracts, full commitment is not feasible as the regulator will be unable to commit to not recontracting on the basis of revealed information. Short-term contracts, in which the regulator bases the next period's mechanism on all available information at the time of contracting, are all that can be expected.

The regulatory contract in the RPI-X game has elements which are incomplete - in the sense that all contingencies are not captured - while other elements, notably the need to share savings with consumers, are 'complete' in the sense that they are not necessarily affected by changes in the firm's operating environment[67]. Game theory therefore suggests that it is optimal for the regulator to commit to those elements of the contract which are 'complete'. Short-term contracts - incorporating this fixed element - allow the regulator to continue to revise the 'incomplete' elements of the contract. This idea that the contract is revised at known periods, but that it incorporates a fixed adjustment mechanism, is consistent with the contributions, discussed by Laffont and Tirole (1993), which allow for 'various degrees of commitment' between the extremes of full commitment and no commitment[68]. It also meets with the National Audit Office's (2002) suggestion that 'the key ingredient in the price control regime should be predictability, not absolute certainty'.

Commitment to the sharing rule increases the level of technical efficiency. The loss of flexibility in the sharing rule will mean, however, that the regulator cannot ensure, at each periodic review, that the current regulatory contract is welfare-maximising given his current information set. In particular, demand and cost conditions are stochastic in the regulated network utility sectors and the regulator will not be able to adjust the sharing rule in reaction to exogenous shocks. This is the cost of obtaining improved technical efficiency and reflects the standard problem with optimal long-term contracts which are ex-ante efficient but ex-post inefficient. Finally we note that the use of a fixed sharing rule, which is essentially an automatic adjustment mechanism, may reduce the

[66]This is also emphasised by Caillaud, Guesnerie, Rey and Tirole (1988) who argue that 'Commitment is (weakly) desirable since the planner can commit to the strategy that it would choose in the absence of commitment'.

[67]Laffont and Tirole (1993) argue that incomplete contracts arise in the context of regulation because regulators 'are unable to set the exact characteristics of the goods to be produced tomorrow or because they may learn tomorrow technological information that cannot be described today'.

[68]The main contributions to the analysis of optimal regulatory contracts in a dynamic environment with varying degrees of commitment are provided by Baron (1989), Baron and Besanko (1984 and 1987), Laffont and Tirole (1988 and 1990) and Lewis and Sappington (1997). Armstrong and Sappington (2002) and Laffont and Tirole (1993, Chs 9 & 10) present useful summaries of the arguments made in these seminal papers.

costs of regulation. There is one less variable for the regulator to decide on at each review, and one less issue for interest groups to debate.

Ideally, with known costs and demand conditions, the fixed sharing rule would be chosen to maximise long-run welfare or to minimise long-run prices[69]. A theoretical analysis of the optimal regulatory lag is provided in Armstrong, Rees and Vickers (1995). They suggest that the optimal lag, and hence presumably our sharing rule, will depend on the firm's current cost level, the elasticity of demand, the marginal cost and marginal benefit of effort to the firm, and the precise relationship between effort and the firm's cost function. The regulator is unlikely to have the required information on these parameters to determine the optimal rule[70]. We therefore expect that the regulator will be required to continue to choose an arbitrary value for the fixed rule.

The arbitrary fixed sharing rule will continue to provide the technical efficiency improvements discussed here, but it will not necessarily improve allocative efficiency as the rule will not maximise constrained welfare. The regulator will be unable to take account of the firm's stochastic operating environment when setting the fixed rule and hence welfare may not be maximised ex-post. In addition, by fixing the sharing rule, the regulator loses the ability to use the sharing instrument as a tool for adjusting the price cap in response to exogenous shocks. For example, the regulator can not reduce the level of savings shared so as to provide the firm with some compensation in the event of a severe negative shock. Similarly, the regulator can not build insurance into the fixed rule to provide some protection to the firm because the impact and probability of the risks are uncertain. We consider, in our next proposal, whether a relative rewards scheme would provide a preferable solution to the technical efficiency incentives problem, given the regulator's asymmetric information and uncertainty about the firm's operating environment.

Proposal 2: A relative rewards scheme

Our second proposal is that the regulator should alter the mechanism so that average industry effort, rather than firm-specific effort, is shared with consumers[71]. The firm's rent is then determined on the basis of its performance relative to the industry rather than relative to the assumptions built into the regulatory contract. Such a sharing rule would lead to an outcome which is similar to that which arises in a competitive market

[69] Armstrong, Rees and Vickers (1995) argue that 'lower prices are closer to expected marginal costs, they induce greater cost-reducing efforts, and they are distributionally beneficial if consumer interests carry more weight than profit in the welfare criterion'.

[70] This is noted by Frontier Economics (2003b): 'In practice, it is not possible for the regulator to identify what this optimal retention period might be'. They argue that this is because information is not available on exactly how the firm's effort choice changes with changes in the regulatory lag.

[71] This proposed amendment to the RPI-X mechanism is only feasible in the water and electricity distribution sectors where more than one firm operates.

- consumers benefit from improved efficiency in the industry through lower prices, and a firm which is more efficient than the average is able to earn above normal returns[72]. This promise of above normal returns provides the firm with an on-going incentive to improve technical efficiency. The proposal is also consistent with the schemes proposed in the literature on optimal regulatory contracts. These schemes pay different rewards to different cost types and do not assume that all firms in an industry earn similar rents.

Under this proposal, the price caps in our two-period operating cost model are:

$$\begin{aligned} \widehat{p}^1 &= \theta - \widehat{e}^1 \\ \widehat{p}^2 &= \theta - \overline{e}^1 - \widehat{e}^2 \end{aligned}$$

where $\overline{e}^1$is the average level of effort in the industry:

$$\overline{e}^1 = \frac{\sum_{k=1}^{n} e_k^1}{n} = \text{ for n firms in the industry}$$

The firm's profit, given these price caps, is:

$$\pi = \left(e^1 - \widehat{e}^1\right) q^1 - \psi\left(e^1\right) + \beta\left[(e^1 - \overline{e}^1)q^2 + (e^2 - \widehat{e}^2)q^2 - \psi\left(e^2\right)\right]$$

The greater the difference between the firm's effort level and the industry average, the higher is the rent earned by the firm. Of course, if the firm's effort level is lower than the industry average it may make a loss.

The profit-maximising effort level is determined by:

$$\begin{aligned} \frac{\partial \pi}{\partial e^1} &= q^1 - \psi'\left(e^1\right) + \beta\left[\left(1 - \frac{1}{n}\right) q^2\right] = 0 \\ &\Rightarrow \psi'\left(e^1\right) = q^1 + \beta q^2 \left(\frac{n-1}{n}\right) \end{aligned}$$

If $\left(\frac{n-1}{n}\right) > (1 - \lambda_o)$ this effort level is strictly higher than that which the firm chooses under the current sharing regime. The larger the number of firms in the industry the closer is the firm's effort choice to the welfare-maximising level. The impact of the proposal may therefore be limited in practice because there are only twelve (ten) companies in the electricity distribution (water) sector. The regulator would be better off calculating the average effort level excluding the firm's own effort level:

$$\overline{e}^1 = \frac{\sum_{k=1}^{n-1} e_k^1}{n-1}$$

[72]Beesely (1994) notes that this reflects the Schumpeterian concept of competition - a dynamic process - rather than the static neoclassical definition: 'The regulator is playing both the role of creating the possibility of earning innovatory gains and that of the 'perennial gale' of competition which tends to then wash away over time'.

With this benchmark, the firm will choose the welfare-maximising level of effort.

The firm has an incentive to undertake high levels of effort because the benefit of any extra unit of savings, relative to the industry average, will all accrue to the firm. This incentive will hold for all the regulated firms, and hence we expect the industry level of effort to increase. This reduces the size of the firm's profit - ie, the gap between the firm's own effort level and that of the industry average will not be very large. We therefore see that the sharing scheme based on relative performance can reduce the rent earned by the firm, but at the same time it provides strong cost saving incentives. That is, allocative and technical efficiency may both be improved.

A further potential benefit from using this relative rewards scheme is that it may reduce the firm's incentive to bias forecasts, as forecast costs no longer impact on the long-run profit stream. Indeed, as noted by Mayer (1999), the relative reward scheme eliminates the need for concern about accurately forecasting costs. This may even mean that the regulator does not need to collect forecasts in the first place, thereby reducing the costs of regulation significantly. We conclude that a sharing rule which is based on performance relative to other firms, rather than performance relative to the regulator's contract, may yield a higher level of welfare.

Commitment to this relative rewards scheme is required to ensure that the firm's decisions are based on certainty about future rules. That is, the regulator must announce ex-ante that a relative sharing scheme will be used, and the rule must not be changed ex-post. The issues raised about commitment in Proposal 1 must therefore also be addressed here. Commitment may be easier with this rule because average industry profits are reduced. In addition, concerns about distribution between consumers on aggregate and overall industry profits will be lessened, and the exogenous pressures on the regulator to change the sharing rule will therefore be reduced.

Furthermore, the regulator does not have to calculate a specific sharing factor. The problems associated with the information requirements of proposal 1 do not arise here. However, several other information issues arise because of the need to ensure that the average effort level of the industry is an appropriate benchmark to judge the regulated firm's effort level against. In particular the regulator must understand, and make adjustments for, any differences in the operating conditions of the firms and any differences in their historical efficiency levels.

The regulator must also consider whether exogenous stochastic shocks affect firms symmetrically or asymmetrically. If the effect is symmetric - e.g. all electricity distribution companies face the same storm and suffer similar consequences - then the yardstick mechanism remains appropriate as the average efficiency level of the industry is a suitable benchmark for an individual firm. However, the shocks may be asymmetric, so that only a subset of firms face the shock or the ability to deal with the shock varies by firm. In these circumstances some firms may be unable to attain the average efficiency level of the industry because of the variation in the distribution and impact of exogenous shocks. The yardstick mechanism may place undue risks on the firm in these circumstances and may limit investments, output delivery and ultimately both technical and allocative efficiency. It is therefore important that the role of the mechanism, and its specific design, take account of the stochastic nature of the sector being regulated. The scheme in practice is likely to be more complex than that described here and the benefits in terms of improved allocative and technical efficiency may be constrained by the extent to which firms' operating conditions are different.

The idea of basing the sharing rule on the industry average performance is not new. It is a direct application of the principle of yardstick competition[73],[74]. The amended sharing rule is also consistent with the spirit of relative price regulation proposed by Mayer (1999)[75]. The proposal is only feasible if the regulator has consistent information on the firms' costs, and if the costs are highly correlated across the industry[76]. In addition, the regulator must ensure that the firms do not collude so as to provide adjusted information which operates in the industry's favour. Finally, the regulator must be willing to allow some companies to suffer a loss when their effort level is significantly below average[77]. It may be difficult for the regulator to commit to this, given the duty to ensure that the firm can always finance its functions, and a regime with less downside risk may be required in practice.

These obstacles are often used as explanations of why yardstick regulation is not observed in practice. For example, Laffont and Tirole (1993) argue that 'relative performance evaluation has not been used much in regulation' because 'regulated firms are often not comparable. That is, idiosyncrasies often prevail over common features'.

[73]See Armstrong, Cowan and Vickers (1994), Armstrong and Sappington (2002), Demski and Sappington (1984), Frontier Economics (2003b), Helm and Yarrow (1988), Mayers and Vickers (1996), Sappington (2000), Sawkins (2001), Shleifer (1985), Sobel (1999), Vickers and Yarrow (1988a), Weyman-Jones (1995) and Yarrow (1989) for detailed discussions on yardstick regulation. Holmström (1982) presents a more theoretical discussion of the value of relative performance measurement in contract design.

[74]Formal yardstick measures are emerging in the regulation of electricity markets in other countries. The reader is referred to Waddams Price (2000a) for a description of the approaches which are currently being developed in Norway and the Netherlands. Waddams Price (2002) also provides a detailed survey of the different approaches used by the UK utility regulators to compare firms' costs within a sector and to determine efficient cost benchmarks. The role of cost comparisons in the electricity sector is also discussed in Jamasb and Pollitt (2000).

[75]The principle of basing the average industry price on the average return in the industry was also discussed much earlier by Yarrow (1989).

[76]Bennett and Waddams Price (2002) emphasise that 'care must be taken that only risks beyond the control of the firms are linked in this way, as weakening a direct link between effort and outcome will result in dampening the incentive structure'.

[77]This is emphasised by Shleifer (1985) who argues that 'It is essential for the regulator to commit himself not to pay attention to the firms' complaints and to be prepared to let the firms go bankrupt if they choose inefficient cost levels'.

We argue, however, that while it may be impossible to design a 'first-best' relative scheme, regulators could attempt to design a 'second-best' scheme. This is emphasised by Helm and Yarrow (1988) who argue that 'Some additional information will always be better than none, and only if market conditions are *entirely* idiosyncratic (which is unlikely to be the case in practice) will the yardsticks effectively be useless'. Yarrow (1989) similarly argues that 'the fact that perfect measurement is impossible does not in any way imply that regulatory systems should not attempt to make the best possible use of the information that *is* available'. The regulator would need to use all available information on the industry when designing the relative scheme. This would enable him to take account of the expected differences across companies - both currently and in the face of uncertain shocks - and would increase the probability of the scheme improving technical efficiency relative to the current regime, while reducing the risks placed on individual firms. Ofwat's approach to assessing operating cost efficiency based on econometric comparisons provides one example of how asymmetric operating conditions might be taken into consideration. Further research is needed on how to take account of the variation in stochastic shocks and on how to translate the comparator analysis into a feasible and welfare-improving relative sharing scheme.

3.2.2 Delivering the efficient quality of service level

In section 3.1.3 we found that the firm may have an incentive to deliver quality of service levels which are below the targets set in the regulatory contract. In addition, even if the targets are met, it is not clear that the level of quality delivered is efficient. This is because the targets, combined with the implied price caps, do not necessarily reflect consumer preferences about the price-quality trade-off. Any solution to the problem of service delivery will require these two issues to be addressed.

We assume that health, safety and environmental targets are fixed, and that there is no scope to change the way in which they are regulated[78]. The regulator does, however, have flexibility over the quality of service target to set for the firm. We consider, in Proposal 3, a way in which the regulator may be able to bring the quality target closer to the efficient level. Proposal 4 discusses how the RPI-X mechanism could be used to ensure that the firm has an incentive to meet this target.

[78]Research into the advantages and disadvantages of market-based regulatory mechanisms - rather than command and control - would be beneficial in this area.

Proposal 3: Offer price-quality menus to consumer groups

We propose that the regulator offers the local consumer group a menu of different price-quality options[79]:

$$\left\{\left(\vec{p}_1^{\,j},\vec{s}_1^{\,j}\right),\left(\vec{p}_2^{\,j},\vec{s}_2^{\,j}\right),\left(\vec{p}_3^{\,j},\vec{s}_3^{\,j}\right),\ldots\ldots\left(\vec{p}_l^{\,j},\vec{s}_l^{\,j}\right)\right\}$$

The consumer group chooses its preferred price-quality pair from this menu. The chosen price-quality pair forms the contract which is offered to the firm in the contract agreement game, and the RPI-X game is otherwise played as described in Chapter 2.

Under this proposal the regulator constructs a (discontinuous) quality supply curve, using fixed assumptions about per unit operating and capital costs, potential efficiency improvements, the sharing of past efficiency savings and the delivery of other outputs. The consumer group similarly derives a quality demand curve and determines the point at which the regulator's supply curve and the demand curve intersect. This process yields an estimate of the efficient quality level given current information and preferences.

The regulator's quality of supply curve will be inaccurate in the sense that the cost forecasts on which it is based will be different to those which emerge ex-post. For example, the firm may be able to deliver the quality of service at a lower per unit cost, shifting the actual quality supply curve to the right of that assumed by the regulator. The change in unit costs may also vary with the amount of quality provided and hence the shape of the quality of supply curve may be different to that assumed. The quality of service level is therefore second-best in the sense that ex-post welfare, based on actual costs, is not maximised. This is an inevitable consequence of asymmetric information and forward-looking price cap regulation.

In addition, both the regulator's supply curve and the consumer group's demand curve will be affected by exogenous shocks. As these are stochastic in nature they will not be built into the cost analysis or into the consumers' preference function. The ex-post efficient level of quality of service may therefore be different to that derived ex-ante. For example, if a water company faces a hot dry summer the per unit costs of providing quality of service increase, exacerbated by increased demand, and the supply function changes. At the same time, consumer preferences for quality may change in the face of restricted output levels and associated quality of service problems. Similarly, an electricity distribution company may face higher costs for delivering a given quality of service if the network is affected by a severe storm. The impact may be less severe than is the case in water as there is no additional requirement to invest in increased

[79]See Baron and Myerson (1982), Laffont and Tirole (1993) and Lewis and Sappington (1989) for a more general discussion of the merits of offering a menu of contracts, rather than a single contract, in a regulatory regime.

capacity, the weather-effect is generally for a shorter period of time, and the change in demand may not be as significant. In either case, the existence of the exogenous shock means that it may be more costly for the firm to reach a given target and/or consumer preferences for a given quality of service may change. The target is therefore unlikely to be efficient ex-post and the firm may face increased costs and risks when it is required to continue to meet it. This problem also arises with the targets currently set by the regulator.

Despite these limitations, there are many benefits of this proposal relative to the current quality regulation regime in the RPI-X game. In particular, the quality of service target, even if second best, is expected to be closer to an efficient level than one chosen on the basis of the regulator's judgement about consumer willingness to pay. The main advantages are outlined here.

- The proposal transfers the responsibility for determining willingness-to-pay from the regulator to the consumer group, and the regulator is only required to calculate the cost of improvements. This reduces the cost of regulation.
- The consumer groups have better information, and better incentives, than either the regulator or the firm to choose the appropriate price-quality trade-offs[80],[81]. They will choose the {price,quality} pair which best matches current information on consumer willingness-to-pay for quality. The consumer group will not have complete information on preferences over time and, in particular, will have uncertainty about future potential exogenous shocks. The chosen target will therefore not be optimal, but it may be closer to the efficient level than that chosen by the regulator or the firm as it is based on the best available *current* information set.
- The use of consumer groups allows the regulatory regime to mirror, to some extent, a competitive market. Here individual consumers are offered a range

[80]The main alternative, considered in the literature and in practice, is to delegate the choice to the firm, as the management is likely to have better information than the regulator about consumer preferences. For example, Lewis and Sappington (1992) derived a mechanism which delivered an optimal {price,quality} pair by inducing 'the firm to employ its superior knowledge and ability entirely in the social interest'. The problem is that the profit-maximising level of quality, given the incentives to reduce costs under the price cap, will not necessarily coincide with the level preferred by consumers. This is a standard externality problem, where the producer bases the decision on private benefits and costs rather than taking account of the social benefits and costs.

[81]Spence (1975) also proposed basing the regulation of quality on a menu of price-quality pairs. He suggested that 'the regulatory authority would manage price-quality trade-offs by confronting the firm, on behalf of consumers, with a reaction function that reflects rates of substitution between price and quality on the demand side of the market'. Spence (1975) recognised, however, that the regulator is unlikely to have the information to determine these 'rates of substitution'. We suggest that this information problem can be overcome by getting the consumer groups to choose from amongst the contracts, and thereby reveal the current price-quality trade-off demanded by consumers.

of price-quality options and choose the one which matches their willingness-to-pay criteria (ie, preferences)[82]. In the utility sectors, the firm cannot provide a different quality of service to different consumers within a region. This is because the quality of service is determined on the network and cannot be varied for individual households. The best the firm can do is offer a level of service which meets the average preferences of consumers within its franchise area. Consumer groups are responsible for representing this 'average consumer preference' in the regulatory regime[83].

- The use of a menu of {price,quality} options ensures that the consumer group is aware of the cost of any quality choice. A key criticism, made by regulators and the Commission, of standard consumer surveys is that consumers will always state a preference for a higher quality of service when information about the associated price is not clarified[84]. The regulator's menu, which incorporates all other assumptions in the allowed revenue calculation, provides the consumers with the information required to compare the set of feasible contracts.

- The regulator offers the consumer group a broad range of options in each period, allowing for the possibility that preferences change over time[85]. The regulator is able to update the information used to construct the {price,quality} menu at each periodic review. A key advantage here is that the regulator can allow consumers to choose whether they want cost savings to be shared as lower prices or as higher levels of service. This is a decision which the regulator currently makes on the basis of his own judgement. Changes in the way in which savings are treated may affect the firm's technical efficiency incentives. The regulator will need to ensure that the impact of the consumer group's choice, between lower prices or higher quality, is revenue neutral from the firm's perspective. A

[82]Frontier Economics (2003b) note that 'If consumers had a choice of network services in a competitive marketplace, they would be able to select the combination of price and quality that best suited their personal preferences. But they cannot "vote with their feet" against a monopoly supplier'. Under our proposal the consumer group 'votes' on behalf of individual consumers.

[83]Spence (1975) provides another justification for presenting the menu of options to the consumer group rather than individual consumers. He finds that the welfare-maximising level of quality is determined at the point where the average benefit of quality equals the marginal cost. Marginal consumer preferences do not provide 'an accurate measure of the social benefits of the increase in quality' because 'in many cases, the marginal consumer is quite unlikely to be representative in his marginal valuation of quality'. For this reason the quality level should be determined using information on the average benefit across all consumers.

[84]Waddams Price, Brigham and Fitzgerald (2002, pp5-6) provide an interesting discussion of the factors which affect consumer responses in willingness to pay surveys.

[85]Forsyth (1999) argues that the 'optimal level of quality will change over time, as incomes rise and customers are prepared to pay for a higher standard of quality, and as the cost of achieving a given level of quality changes with technological developments'. In our scheme, consumer groups are able to express these changes in their contract choices at each periodic review.

fixed sharing rule will help here. If the firm retains the same proportion of the savings made, and the consumer group determines how the fixed consumer share is divided between price reductions and quality improvement, there should be no impact on the firm's cost saving incentives.

We conclude that the regulated industry could move closer to an efficient level of quality of service if the consumer group chose the preferred price-quality pair from a menu offered by the regulator. The improvement, relative to the current system where the regulator chooses the price-quality pair, arises because the consumer group is assumed to have better information on current preferences. The target may not be optimal ex-post, however, as both the demand and supply curves will adjust in the face of cost efficiency improvements and exogenous shocks. This is therefore a move towards the optimum rather than a mechanism which reaches the optimum with certainty.

The proposal is relatively simple to implement and would not increase the costs of the regulatory regime significantly. The regulators currently use financial models to determine the size of the price cap under a number of different assumptions. These models could be used to calculate a range of different prices for different quality levels, holding all other factors constant. Similarly, the regulator already meets with consumer groups during the periodic review and discusses the price-quality trade-off issue with them. The menu of {price,quality} options would provide a clear focal point on which to base these discussions, but would not increase the level of interaction required.

The area in which costs will increase is for the consumer groups, who will be under pressure to ensure that they have up-to-date information on consumer willingness-to-pay for quality of service. The consumer organisations would therefore need to be provided with adequate resources to ensure that they can collect the required information on consumer preferences. The powers and duties of the consumer groups would also need to be clarified in the legislation, and brought in line with those of the regulator, the firm and the Commission. Markou and Waddams Price (1999) stress, in this regard, that care should be taken to ensure that consumer groups are required to consider both the costs and benefits of any decision. If only the benefits are considered there is a danger that consumer groups will 'become real champions of a particular cause', irrespective of the impact on investment costs[86]. In addition, the independence of the consumer groups would need to be ensured.

[86] Markou and Waddams Price (1999).

Proposal 4: Devising a penalty for not meeting the target

When Proposal 3 is implemented, the quality of service target will be the best available current estimate of the efficient standard. In this case, both overshooting and undershooting the target are inefficient. The regulator will want to penalise the firm for underperforming relative to the contract, and will not want to provide an incentive for the firm to outperform the target. The consumer group has indicated that it does not require a higher level of service and, in particular, that it is not willing to pay a higher price for the higher level of service[87].

We propose that the current financial incentive mechanisms - namely compensation payments and the price cap adjustment - be removed, and replaced with an ex-post penalty. We assume that, as now, the firm's performance will always be audited and that the penalty will be imposed if a breach of the target is identified[88]. The regulator must show that the ex-post penalty is credible by including the details of the scheme, and the penalty size, in legislation. Without such credible commitment, the firm will assume that there is a positive probability that the penalty will not be used, and it may risk reducing the quality of service below the target.

To determine the size of the penalty the regulator must consider the firm's incentive compatibility constraint[89]. In the absence of any penalty regime the firm will deliver a lower level of service if:

$$\Delta\pi^c\left(s_l^j\right) + H\left(s_l^j\right) > H\left(\widehat{s}^j\right)$$

The regulator should therefore set a penalty, F^S, which makes the firm indifferent between choosing s_l^j and $\widehat{s}^j$. As the penalty can only be imposed after the level of service has been chosen, it will be discounted by the firm. The regulator must therefore choose the penalty such that:

$$\begin{aligned}\Delta\pi^c\left(s_l^j\right) + H\left(s_l^j\right) - \beta F^S &= H\left(\widehat{s}^j\right) \\ &\Rightarrow \ F^S = \left(\frac{1}{\beta}\right)\left(\Delta\pi^c\left(s_l^j\right) + H\left(s_l^j\right) - H\left(\widehat{s}^j\right)\right)\end{aligned}$$

[87] If the firm delivers a higher level of service, with the same or a lower level of costs than assumed, the regulator will use the revealed information to offer new price-quality pairs to the consumer in the next period. In this way consumers do have the option of getting higher quality at the same price in the next period, but they are not required to pay for a higher quality of service which they did not request.

[88] If auditing is costly the regulator would need to trade-off the cost of monitoring the firm's performance against the benefit of being able to punish the firm for breaching the standard set. This trade-off may affect the design of the optimal penalty regime. Further analysis of mechanism design with costly auditing can be found in Baron and Besanko (1984), Laffont and Martimort (2002) and Laffont and Tirole (1993, Ch12).

[89] Waddams Price, Brigham and Fitzgerald (2002) suggest that the penalty should be based on the value of the service lost to consumers. This is very difficult to measure, particularly when consumers themselves may not be able to place a value on the impact of any reduction in service quality.

The penalty could be incorporated into the allowed revenue calculation at the periodic review. This ensures that the firm's consumers automatically get compensated. The allowed revenue formula, for year one of the regulatory period, becomes:

$$\widehat{R}_1^j = \widehat{\theta}_1^j q_1^j + \widehat{\delta}_1^j + \left(\widehat{r}_1^j \times \widehat{RCV}_1^j\right) - F^S$$

The key difference between this proposal and the existing price cap adjustment is that the firm is only penalised if it underdelivers relative to the target chosen by its own consumers, rather than being judged relative to other firms in the industry.

In practice, the regulator is unlikely to have the information required to set this fixed penalty exactly. The reputation value will not be quantifiable, and the regulator will not, as emphasised elsewhere, be able to identify the proportion of profits earned which are attributable to underinvesting in quality of service. The best available alternative is to simply set a very high, arbitrary, penalty. If the firm believes that the regulator is committed to the penalty, it will not risk choosing a lower level of service and, hence, the penalty will never be charged. The regulator's ability to commit may, however, be restricted by its requirement to ensure that the firm can finance its functions. A very high penalty, which might result in the firm making a financial loss, would breach this duty and, hence, the firm would assume that the regulator cannot credibly commit to it. It may therefore be easier for the regulator to commit to a penalty which was set equal to a fixed proportion of the firm's profits (eg, 10%), as this would always ensure that the firm retained some profits after the penalty was imposed.

The regulator must exercise caution when imposing the high penalty on the firm. Ex-post if may be more difficult, and more costly, for the firm to reach the required target given exogenous shocks in its operating environment. The regulator should therefore only impose a high penalty when the failure to reach a target is attributable to firm behaviour rather than exogenous shocks. It is difficult to make such distinctions however. This increases the complexity of the regime but reduces the risks that the firm faces. It also creates a further problem, whereby the firm may attempt to influence the information provided to the regulator, and the interpretation of that information, to reduce the probability of facing a penalty.

Given these problems with the penalty regime, it may be beneficial to retain the relative price cap adjustment, allowing the regulator to use information on other firms' performance as a check. For example, if all firms failed to meet their target then the regulator may conclude that there is no need for an adjustment because they were all affected by an exogenous shock. If, however, one firm failed to meet its target but all other firms met theirs then that firm could face a downward price cap adjustment.

The regulator must ensure that exogenous shocks affect firms symmetrically before using industry performance as a benchmark.

The price cap adjustment may also be preferable to the large fixed penalty because the regulator can moderate the size of the downward adjustment so as to take account of the risk of exogenous shocks affecting firm performance. The flexibility in the price cap adjustment may therefore be beneficial. We continue to suggest, however, that the adjustment should only be downward as there is no case for rewarding the firm for producing more quality than is demanded by consumers. Only if the regulator found that exogenous shocks led to an increase in quality of service demanded during the period would such outperformance be justified.

3.3 Conclusion

RPI-X regulation was chosen for the utility sectors in the UK because, as a fixed price cap regime, it was expected to reduce monopoly power, and thereby increase welfare, by providing the firm with an incentive to reduce costs. We found, in Chapter 2, that the regime which operates in practice is different from that assumed in the theory of pure price cap regulation. It is therefore, perhaps, not surprising that the welfare properties of the actual regime are more complicated than originally envisioned.

The main problems, discussed in detail in section 3.1, are summarised here.

- Allocative efficiency is not delivered because of asymmetric information about the firm's future costs, and because the technical efficiency objective requires the firm to be able to earn a positive rent. In addition, welfare is not maximised when the regulator is unable to pay transfers to the firm. These problems arise with all mechanisms designed under the same conditions. The allocative inefficiency problem is exacerbated in the RPI-X game because the rules used to share cost savings with consumers are not chosen to maximise total welfare.

- Cost savings incentives are delivered, but they are not as strong as expected in the theory of pure price cap regulation. This is because the regulator shares savings with consumers over time. In addition, the sharing rule is chosen arbitrarily and, hence, the firm makes its effort decision in the face of uncertainty about the marginal benefit of each unit of effort. A ratchet effect problem also arises because the regulator bases the price cap on historic cost information. Finally, the RPI-X mechanism distorts the firm's input choices and thereby reduces technical efficiency.

- The final element of welfare which we considered is the efficient delivery of outputs. We conclude that the RPI-X regime provides the firm with some protection against the hold-up problem, but the incentive to reduce costs may lead to underinvestment in output delivery. The existing output regulatory regime may not be sufficiently strong to offset this effect. In addition, the regime is focused on standards which do not reflect consumer willingness-to-pay and, hence, it is not clear that the quality of service provided is efficient.

Our conclusion, therefore, is that while welfare may be higher under the current RPI-X mechanism than would be the case with no regulation, it is not maximised. The problems are driven by the legal and institutional framework within which decisions are made - ie, the characteristics of the RPI-X game - and by the way in which the regulator chooses to set the price cap and output targets. It is difficult to change the characteristics of the game, but the regulator has the freedom to change the methodology used to design the regulatory contract.

This suggests that regulators, and other interested parties, should consider ways of changing the RPI-X mechanism, *taking account of the characteristics of the existing game*, so as to increase the level of welfare. We have proposed a small number of possible changes in section 3.2. These suggest that allocative efficiency might be increased by using relative rewards schemes. Technical efficiency incentives might be improved by fixing the sharing rules used over time and/or by basing the sharing rule on a relative scheme. Finally, the efficient level of service might be delivered by allowing consumer groups to choose their preferred price-quality combination from a menu of contracts, and by imposing an ex-post penalty on the firm for underdelivering relative to this chosen quality level.

The actual improvements delivered by these proposals will depend on the characteristics of the regulated sector. In particular, asymmetric information, a stochastic operating environment, and significant variation across firms may limit the feasibility and welfare-improving potential of the alternative mechanisms. Further research is required to identify how these factors constrain the potential improvements in allocative efficiency, technical efficiency and quality of service delivery.

These proposals are preliminary at this stage. An on-going research programme should include a theoretical analysis of the precise net welfare effects of each proposal[90]. In particular, we would want to examine the impact of each proposal on the other proposals, and on the existing elements of the RPI-X mechanism. For example, the technical efficiency gains arising from the regulator committing to a fixed sharing rule should be weighed against the potential allocative efficiency costs of not being able to adjust the contract in reaction to changes in the stochastic operating environment.

[90]The detailed analysis required to determine the exact properties of our proposed changes reflects Helm's (1995a) warning that 'To identify faults in the current regulatory regime is however relatively easy. To design a system which improves upon it is much harder'.

This welfare analysis should take account, in the first instance, of the constraints imposed by the fixed characteristics of the RPI-X game and the nature of cost and demand conditions in the regulated sector. We should also, however, consider whether the proposals could be improved if some of these characteristics, such as the regulator's objective function, were changed.

A feasibility study - presumably at sector level - also needs to be undertaken. Evidence on regulatory mechanisms in other countries could provide useful insight on what type of adjustments do and do not work. For example, profit-sharing schemes in the US telecoms sector could be examined to determine how a fixed sharing rule might be calculated. In Latin America, several regulatory agencies operate under detailed licence conditions which explicitly outline the methodologies used to set the regulatory contract. These case studies could be used to evaluate the impact of commitment devices on welfare. Any comparisons, across countries and/or sectors, need to take full account of differences in the regulatory regimes which are in place. This is because, as we have learned from our analysis of the RPI-X game, the way in which any regulatory mechanism is designed and implemented has significant implications for its actual impact on welfare.

Chapter 4

Productivity growth in the water and electricity sectors

RPI-X regulation was introduced in the UK utility sectors on the grounds that it outperformed other regulatory regimes, notably rate of return regulation, in terms of its ability to control a firm's monopoly power. In particular it was argued that price cap regulation provides the regulated firm with an incentive to reduce its costs over time. In this chapter we examine whether this promise has been delivered by calculating productivity growth rates in the water and electricity sectors in England and Wales. Chapter 3 discusses the productivity improvement incentives from a more theoretical perspective.

Annual productivity growth rates are calculated for the following firms.

- The ten *water and sewerage companies* (WASCs) - the companies which hold a licence to provide monopoly water and sewerage services within a specified region in England and Wales.

- The *distribution businesses of the twelve regional electricity companies* (RECs) in England and Wales - the companies which hold a monopoly licence to distribute electricity from the high-voltage network, along lower-voltage power lines, to the end-user within a specified franchise area[1].

- The *transmission business of the National Grid Company* (NGC) - the company which holds the monopoly licence to transmit electricity from upstream generation plants, along high-voltage lines, to the distribution network. The licence area covers all of England and Wales.

[1]These twelve companies are also referred to as Public Electricity Suppliers (PESs), along with ScottishPower, Scottish Hydro-Electric and Northern Ireland Electricity.

These companies have been subject to RPI-X regulation since 1990. We calculate productivity growth for the period 1991 to 2000, covering at least two regulatory periods for each of the sectors[2]. We stress that the objective is not to carry out a comparative efficiency analysis of the firms in each sector. Instead we focus on sectoral trends over time. Appendix B provides further information on these companies.

Productivity is measured in two ways in this analysis. First, we examine changes in the *firm's costs*. The regulatory regime is expected to deliver large reductions in unit costs over time. Annual changes in real unit operating costs and real unit capital investment are presented to determine the extent to which the price cap regime has delivered these anticipated cost savings.

Productivity involves more than a reduction in costs however. This is particularly true if cost reduction is associated with a reduction in the amount produced and/or by a reduction in the quality of service provided. We therefore also measure trends in the firms' *technical efficiency level*. There is an improvement in technical efficiency if the regulated firm is producing the same amount of outputs using less inputs, or more output using the same amount of inputs.

We consider trends in technical efficiency using estimates of total factor productivity (TFP). Three alternative measures of TFP are calculated - a Tornqvist Index using calculated input shares, a Tornqvist Index based on econometrically estimated input shares, and a Malmquist Index calculated using data envelopment analysis (DEA)[3]. For our short data set no one methodology is ideal, and we wish to examine the impact of the methodology choice on the productivity assessment. Using a number of methodologies also allows us to provide a range of TFP growth rates. This is considered preferable when individual point growth rates may suffer from potential biases.

Having identified the changes in cost efficiency and technical efficiency since privatisation, we turn to the question of how the RPI-X mechanism has influenced this productivity performance. Our analysis in Chapters 2 and 3 suggests that the price cap setting process affects the firm's effort choice, and hence the extent of productivity improvement. The elements which are of most interest are the expected length of time over which cost savings are retained, and the impact of regulatory commitment on

[2]The regulatory periods over which the price caps were set are as follows:
NGC - 1990 to 1993; 1993 to 1997 and 1997 to 2001
RECs - 1990 to 1995 and 1995 to 2000
WASCs - 1990 to 1995 and 1995 to 2000

[3]Because of the short data set, we are unable to estimate a production function for NGC. We therefore do not calculate Tornqvist Indices with estimated input shares for NGC. One possible proxy measure for NGC's estimated input share is the input share from the electricity distribution production function. It was decided that this proxy was inappropriate given the technological and cost structure differences between the transmission and distribution networks. Similarly, Malmquist Indices, which are calculated using information on comparator companies, could not be calculated for NGC as it is a national monopoly and distribution companies were considered inapproriate comparators.

this expected regulatory lag. We also consider the impact of takeovers and mergers, and the level of quality of supply, on productivity trends during the period. Finally, we introduce a variable which controls for the number of years since privatisation.

The chapter is structured as follows. Section 4.1 summarises related literature. The variables of interest and the data sources used in the analysis are outlined in Section 4.2 and described in detail in Appendix C. Section 4.3 presents information on changes in unit cost levels since privatisation. The different methodologies used to calculate TFP are defined in Section 4.4, and the average annual TFP growth rates are presented in section 4.5. The impact of the regulatory regime on cost efficiency and technical efficiency is examined in section 4.6. Section 4.7 concludes.

4.1 Related literature

A vast literature exists on productivity in the regulated sectors. To ensure that this review is focused and concise we only consider two aspects of the relevant literature here. First, we look at the body of literature which measures the rate of productivity growth in the regulated utility sectors in England and Wales. We contribute to this body of literature by providing updated estimates of productivity growth in the water and electricity sectors, and by using a number of different methodologies to measure productivity. Our research also stands out for incorporating more than one industry, enabling us to undertake cross-sectoral comparisons. The second body of literature which we are interested in analyses the impact of incentive regulation on productivity growth. We contribute to this literature by examining how the price cap setting process in the RPI-X regime affects the productivity growth rate. Our analysis is carried out using a cross-sectoral panel dataset, with all companies in the panel subject to RPI-X regulation since 1990/91.

4.1.1 Productivity in the UK utility sectors

In this section we summarise the literature which measures productivity growth in the regulated utility sectors in England and Wales. The studies on productivity in the electricity and water sectors relate most closely to our work[4]. We first discuss the large volume of research on the impact of privatisation on productivity.

[4]A study of productivity in the gas sector can be found in Waddams Price and Weyman-Jones (1996).

The privatisation effect

A large number of studies have examined the impact of the Conservative Government's privatisation programme on productivity. The impetus for this research was the need to test the claim, from economic theory, that privatisation would lead to an increase in productivity[5].

We do not provide details of individual papers here but instead refer the reader to Europe Economics and Crafts (1998, Ch7) and Waddams Price (2000a) for useful reviews and summaries of the main studies. More recent research and literature surveys can be found in Pollitt (1999) and Green and Haskel (2001). The results of the studies are quite mixed, suggesting that privatisation had an uncertain impact on productivity. Bearing this in mind, the broad conclusions reached are outlined here.

- There was an increase in productivity in most industries after privatisation, although in some cases the growth in total factor productivity was quite low.
- Privatisation had a greater impact on labour productivity than on total factor productivity. This suggests that there may have been substitution across inputs. In particular, reductions in employment costs may have been offset by increases in the capital stock and/or a reduction in turnover.
- Productivity growth was simultaneously affected by the change in ownership, market liberalisation, corporate restructuring, and regulation. It is difficult to separate out the effect of each individual policy change because there is no obvious counterfactual which the evidence can be compared against. Some studies have concluded, however, that the transfer of assets to private ownership had the least impact on productivity growth, while the opening up of previously monopoly markets to competition had the strongest positive effect. Corporate restructuring and competition were also found to have had a greater impact on productivity than regulation in a number of studies. Most of the studies only consider the period up to 1995/96 and hence, as noted by Green and Haskel (2001), 'stop before the effect of regulation can be estimated with any reliability'.
- Productivity improvements did not always coincide with the timing of privatisation. For many companies productivity growth was observed before flotation.

[5]Details of the privatisation programmes in each of the utility sectors can be found in Armstrong, Cowan and Vickers (1994), Newbery (1999), and Vickers and Yarrow (1988a). See Beesley (1997), Markou and Waddams Price (1999), Pollitt (1999), Vickers and Yarrow (1988a), and Yarrow and Jasiñski (1996), for discussions of the expected costs and benefits of privatisation.

In other sectors there was a decline in total factor productivity at privatisation, but pre-privatisation levels were restored after a few years.

Productivity in the electricity distribution sector

In this section we summarise the main findings in the recent literature on productivity growth in the electricity distribution sector in England and Wales[6]. The conclusions which can be drawn from this literature are that productivity improved just before privatisation, and there was a subsequent significant increase in total factor productivity after the 1994 price review.

- *Burns and Weyman-Jones (1996)* found that cost efficiency improved slightly after privatisation. This was deemed to be a 'one-off effect' and, it was claimed that 'privatization has not affected the organization of inputs in the productive process'. The authors also stress that the change in cost efficiency may have been caused by changes in accounting policies rather than the privatisation process per se.

- *Hattori, Jamasb and Pollitt (2002)* found that TFP declined in the electricity distribution sector just before privatisation, but there was a large increase between 1995 and 1996.

- *Tilley and Weyman-Jones (1999)* found that average annual growth in TFP was 6.3% for the period 1990/01 to 1997/98, with the highest growth rates after the 1994 price review. The growth rate varied significantly across companies, with a lowest rate of 3.6% and a highest rate of 9.4%. They also concluded that the productivity growth reflects technological change (ie, a shift in the productivity frontier) rather than efficiency change. This suggests that the regulatory regime has not led to a significant catch-up effect across companies in the sector, but 'the industry as a whole is responding to the technical efficiency incentives of privatisation'.

- *Weyman-Jones (1995)* found that efficiency improved just before privatisation, and the variance in efficiency levels fell across companies. He also emphasised that comparative studies need to take account of differences in the firms' operating environments.

[6]A range of different methodolgies have been adopted in these papers. In some cases regulatory cost functions are esimated using ordinary least squares, generalized least squares, and within-groups. In others, real unit operating cost changes and total factor productivity growth have been calculated directly using accounting data. Finally, some authors, notably Weyman-Jones, have used data envelopment analysis and stochastic frontier analysis to compare efficiency levels within the sector and over time. Further details can be found in the individual papers.

- *Weyman-Jones (2001a)* examined whether Ofgem's relative efficiency rankings in the 1999 distribution price review were sensitive to the type of methodology used to determine the firms' efficiency levels. Ofgem's analysis was based on a simple corrected ordinary least square regression analysis. Weyman-Jones found that the rankings did not vary significantly when other methodologies were used, but the actual efficiency scores did vary slightly. He concluded that 'the use of stochastic and non-convex methods of nonparametric efficiency measurement' are preferable.

- *Weyman-Jones (2001b)* provided a brief summary of the main studies on productivity growth in the electricity distribution sector. The main conclusion reached is that 'In the immediate aftermath of privatisation, productivity growth seemed not to differ markedly from pre-privatisation experience, but, following the 1994 control, considerable improvement can be seen'.

Productivity in the water sector

In this section we summarise the recent literature on productivity growth in the water and sewerage sector in England and Wales[7]. The conclusions from this research differ slightly from those outlined above for electricity distribution. There has been a significant reduction in operating costs in the sector, but this has not been matched by an improvement in total factor productivity. This suggests that there has been substitution from non-capital inputs to capital inputs. It may also reflect the fact that output measurement in these calculations does not always take account of the increases in quality which have been delivered.

- *Europe Economics and Crafts (1998)* found that water companies reduced real unit operating expenditure by 3.8% per annum between 1993 and 1998. This was largely due to an average annual increase in labour productivity of 4.6%. The study finds that the reductions were driven, in equal proportion, by shifts in the sectoral efficiency frontier and by a narrowing of the relative efficiency gap between companies.

- *Hunt and Lynk (1995)* examined the impact of the government's policy, at privatisation, to separate the functions of production, economic and environmental regulation in the water sector. The conclusion reached was that the separation policy reduced the economics of scope in the industry, and that this would lead to

[7] As with the electricity distribution studies, a broad range of methodologies have been used to measure producitivity in the water sector. Attempts have also been made to incorporate improvements in drinking water and environmental quality as outputs in the production process.

an increase in costs. There is no analysis, however, of how the privatisation and regulation process post 1989 may have compensated for this loss in efficiency.

- *Lynk (1993)* examines the impact of private ownership on efficiency by comparing the performance of the private statutory water companies (the current water only companies) to that of the public regional water authorities (the current water and sewerage companies) before privatisation. The study concludes that average inefficiency was substantially higher in the private sector than in the public sector. The author concluded that there is no evidence that private ownership is superior in this sector.

- *Markou and Waddams Price (1999)* found that average annual labour productivity growth was higher before privatisation (7.8%) than afterwards (6.5%). They also emphasised that, according to Ofwat, operating costs fell by 8% in the water sector between 1992/93 and 1997/98, and by 10% in the sewerage sector.

- *Saal and Parker (2001)* found that non-capitalized labour productivity was higher post-privatisation, with the largest improvement in the 1995-99 period[8]. In contrast, there was a decline in average annual TFP growth after privatisation. This suggests that the increase in capital investment overshadowed the reduction in labour costs and any increase in the quality-adjusted output level. The productivity measures used incorporate an output index which reflects 'both the quantity and quality of water and sewerage services'.

- *Shaoul (1997)* used a value added or net output approach to measure efficiency[9]. He found that 'Significant increases in efficiency had occurred prior to privatisation leaving little room to improve efficiency without jeopardising levels of service and future service provision'. The conclusion reached is that ownership change had little effect on efficiency because, according to the author, there was no scope for ongoing improvements in this area.

- *Thanassoulis (2000)* described the results of a data envelopment study commissioned by Ofwat for the 1994 periodic review. The study predicted that operating costs could be reduced by 26.67% during the 1995 to 2000 period. This improvement would arise from inefficient firms catching up with the estimated 1992/93 efficiency frontier, and takes no account of potential shifts in the frontier over

[8]Details of productivity growth rates for individual water and sewerage companies can also be found in this paper. Similar results are presented in Saal and Parker (2000).

[9]Saal and Parker (2001) argue that using value added as an output measure is commonly regarded as 'inappropriate in regulated industries where prices are not determined in markets'.

time. The results from the DEA analysis were similar to those derived by Ofwat using regression analysis.

4.1.2 Impact of regulation on productivity

In this section we discuss the literature which examines the impact of incentive regulation on productivity. The literature is divided into two strands: studies which focus on regulation in the US, and those which consider productivity change in other countries. A discussion on the principles of how to empirically test the impact of regulatory schemes can be found in Joskow and Rose (1989).

The US experience

We summarise a selection of recent studies which discuss the impact of regulation on productivity in the US. These studies generally focus on the state telecommunications markets, where comparisons can be made across different regulatory regimes over time. Kridel, Sappington and Weisman (1996) provide a complementary survey of earlier literature in this area. The general impression which we get from the studies summarised here is that it is questionable that incentive regulatory mechanisms, including price cap regimes, have delivered on their promise of improving productivity significantly relative to rate of return regulation.

- *Ai and Sappington (2002)* found, for the period 1986 to 1999, that the operating costs of the regional Bell operating companies were 4.5% lower with rate case moratoria than with rate of return regulation. In contrast, costs were not lower, on average, with earnings sharing or price cap schemes relative to rate of return regulation. The cost comparison changed when applied to the regulated firms which operated in local competitive markets. In this environment, costs were lower with all three incentive schemes than with rate of return regulation. This suggests that 'local competition and incentive regulation appear to play complementary roles in motivating cost reductions'.

- *Granderson and Linvill (1996)* examined the productivity growth of twenty interstate natural gas pipelines companies which were subject to rate of return regulation between 1977 and 1987. TFP was found to grow at an average annual rate of 6.4%. Regulation was found to have had only a small effect on the level of growth, but it did have an impact on the characteristics of the production technology. When the effects of regulation were taken into account the proportion of productivity growth which was attributed to scale economies was reduced by 38%, and the proportion which was attributed to technical change was increased by 28%.

- *Majumdar (1997)* considered the impact of different types of incentive regulation on productivity in the US local telecoms market between 1988 and 1993. He found that 'while the introduction of a pure price-cap scheme has a positive effect on technical efficiency, the introduction of an earnings-scheme alone eventually has a detrimental effect on technical efficiency'. In addition, it was found that a combined price-cap and earnings-scheme had a positive effect on productivity, but the impact was smaller than with a pure price-cap scheme. This suggests that a pure price cap regime is the best means of improving productivity.

- *Resende (1999)* examined the impact of incentive regulation on productivity growth in the local telecommunications sector. The main conclusion reached was that 'alternative regulatory regimes (price-cap and incentive regulation) do not seem to play any role in improving technical efficiency, in comparison to traditional rate-of-return regulation'. This contrasts, significantly, with the predictions of economic theory.

- *Roycroft (1999)* calculated the TFP growth rates of the local exchange carriers owned by Ameritech for the period 1990 to 1997. The productivity growth rate was then regressed on a number of regulatory and technology control variables to assess the impact of the regulatory regime on productivity. The author found that 'the introduction of price cap and incentive regulation leads to statistically significant increases in TFP growth'.

- *Uri (2002)*[10] measured productivity improvement in the telecommunications sector between 1988 and 1999. The analysis focused on the nineteen local exchange carriers which have operated under price cap regulation since 1991. TFP was found to have grown by about 5.5% per year. The author noted that the 'growth is due primarily to innovation rather than improvements in efficiency'. This suggests that regulation has had little impact on productivity during the period.

Evidence from other countries

In this section we summarise a number of studies which analyse the impact of price cap regulation on productivity in the UK and Spain. The main lesson from these studies is that price cap regulation has led to operating cost reductions, but it is not clear that it has led to an overall improvement in productivity.

[10] A very similar analysis is found in Uri (2001). As the studies, and results, are so similar we do not discuss this second paper here.

- *Arocena and Waddams Price (2001)* examined productivity growth differences between the publicly and privately owned Spanish coal-fired generation companies in the period 1984 to 1997. All the firms, independent of ownership type, were subject to incentive based regulation since 1988. Cost-plus regulation had been used before then. The authors found that public sector firms were more efficient before the price cap mechanism was introduced. The efficiency of the private firms increased when price cap regulation was introduced. Much of this increase occurred at the time that the regime was changed, with little further improvement under the price cap system. In addition, the average improvement was driven by the less efficient firms catching up with the frontier, while there was only 'modest' change in the frontier itself.

- *Domah and Pollitt (2001)* undertook a social cost-benefit analysis to assess the impact of privatisation and restructuring on the UK electricity supply and distribution businesses between 1986/87 and 1997/98. They found that 'there was a rise in real unit distribution and supply controllable costs by about 15 per cent immediately after privatisation in 1990. The cost remained at a high level until 1994-95, after which there was a dramatic fall'. The largest fall came after 1996/97. This trend demonstrates, according to the authors, 'a clear relationship between the cost changes and changes in the regulatory 'environment''. In particular, the cost falls coincided with the tight distribution price controls set for the 1995 to 2000 period.

- *Hattori, Jamasb and Pollitt (2002)* compared the cost efficiency of the distribution companies in England and Wales to that of Japanese electricity distribution companies which are subject to rate-of-return regulation. They found that the average efficiency scores 'are declining over time in both countries' and that the 'UK electricity distribution shows significantly better performance over the Japanese electricity distribution' between 1995/96 and 1997/98. The authors used these results to tentatively conclude that 'the relative changes over time seem to indicate the effectiveness of incentive regulation'.

- *Saal and Parker (2001)* examined the question of whether the tightening of the price cap in the 1995 water periodic review led to an improvement in productivity growth. They found that non-capitalised labour productivity growth increased significantly between 1995 and 2000, but there was no significant increase between 1985-90 and 1990-95. The authors concluded, from this, that 'the more rigorous economic regulation embodied in the 1994/95 price review was the primary stimulus for operational efficiency gains and not the change of ownership per se'. The tightening of the regulatory regime does not appear to have had a similar impact on TFP growth, with the average annual growth rate declining in the 1995-2000 period relative to the early years after privatisation. The authors therefore concluded that 'the productivity results are not consistent with the hypothesis that the regulatory system became more effective in generating efficiency gains after the price review'.

- *Tilley and Weyman-Jones (1999)* used a panel data regression to examine the factors, including the timing of the price cap reviews, which may have affected productivity growth in the electricity distribution sector since privatisation. They found that the TFP growth rate was affected by a structural break at the time of the 1995 distribution price control. There is no other explicit analysis of the impact of the regulatory regime on the productivity growth rate however.

4.2 Relevant variables

Details of the variables which are used to calculate productivity, and the relevant data sources, are provided in Appendix C and summarised here. Financial data is taken from the firms' annual regulated accounts and non-financial data is taken from reports published by the regulators and the Centre for Regulated Industries.

Output is measured as the volume delivered in the electricity sectors, and the number of properties connected to the network in the water sector. We also use a quality-adjusted output measure. This takes account of changes in the proportion of output which is of 'good quality' over time. In the electricity sector, quality relates to the level of service interruption in the sector. In the water sector, a weighted quality index is used which captures the quality of service provided to consumers in the water and sewerage sectors (ie, water pressure and sewerage flooding problems), drinking water quality and environmental quality. Turnover is used to measure the value of output to the firm.

We assume that two inputs are used in production, capital and non-capital. The capital level is measured as the current cost value of gross fixed assets at the year-end. The capital cost is calculated as the sum of the financing cost of capital plus the cost of depreciation:

$$(\text{Rental rate} \times \text{Capital level}) + \text{capital charges}$$

We also consider the level of gross capital expenditure when examining trends in unit costs. This is equal to the annual additions to fixed assets.

The level, and value, of non-capital inputs is measured as the value of the firm's controllable operating costs. The financial measure is used because there is no clear physical measure which encompasses all the non-capital inputs.

4.3 Cost efficiency

The primary theoretical benefit of RPI-X regulation is that it provides the regulated firm with an incentive to reduce its costs. We wish to determine the extent to which this is true by examining the annual rate of change in real unit costs for the regulated businesses.

4.3.1 Real unit operating costs

We first consider the rate of change in real unit controllable operating costs (RUOC). This measure allows us to compare the performance of different firms while taking account of differences in the scale of activities. We look at the standard measure of unit operating costs and a measure which takes account of quality of supply:

$$
\begin{aligned}
RUOC &= \frac{\text{Controllable operating Costs}}{\text{Output}} \\
QRUOC &= \frac{\text{Controllable operating Costs}}{\text{Quality-adjusted output}}
\end{aligned}
$$

Figure 4.1 shows the trend in $RUOC$ since privatisation and Figure 4.2 shows the trend in $QRUOC$ since privatisation. We see that there has been a large decrease in real unit operating costs in all sectors since 1990. The trend in the annual rate of change is not significantly altered when we use the quality-adjusted output measure. The rate of decline has been smoothest in the electricity distribution industry. There was a step-jump in the water industry in 1992/93, which was followed by a smooth and small decline. This was driven by the large reduction in employment costs for all water companies in 1992/93[11]. NGC has had the largest overall decline.Table 4.1

shows the average annual change in real unit operating costs for each of the sectors. Annual change is measured as the log level in year t minus the log level in year t-1. Again, we see that there have been large reductions in real unit operating costs in all sectors since privatisation. This suggests that the cost reduction incentive properties of the regime are working. In the electricity sector, the rate of decline was highest in the 1995-2000 period. In contrast, it was higher immediately after privatisation in the

[11]The large reduction in employment costs reflects a reduction in the number of employees in the sector. We expect, however, that it also represents a transition from direct employment contracts to contracting-out, which should be picked up in the controllable operating costs measure. Furthermore, companies may have reduced the level of employment costs classified as direct operating costs but increased the level of capitalised employment costs. Accounting changes such as this emphasise the need to use total cost measures rather than operating costs only to examine trends in productivity.

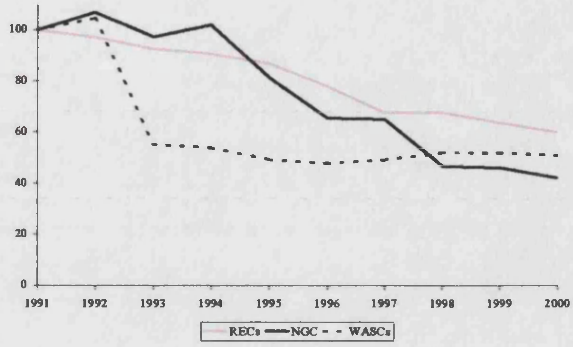

Figure 4.1: RUOC Index

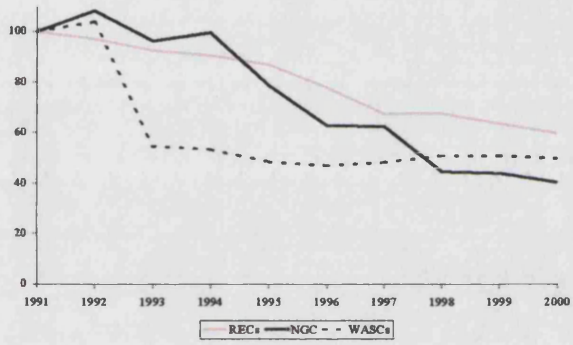

Figure 4.2: QRUOC Index

water sector, reflecting the large reduction in labour costs in 1992/93. These results are consistent with the literature discussed in section 4.1. The rates of decline change marginally when we use the quality-adjusted output measure.

Table 4.1: Average annual change in RUOC

	1990-95	1995-2000	1991-2000
RUOC			
Distribution	-3.73	-7.74	-7.75
NGC	-5.18	-13.2	-8.65
Water	-18.06	0.65	-7.67
QRUOC			
Distribution	-3.75	-7.75	-7.98
NGC	-6.01	-13.4	-8.99
Water	-18.4	0.57	-7.87

4.3.2 Real unit capital costs

We also consider the firm's performance with respect to real unit capital costs (RUCC). The measures used for unit capital costs are:

$$\begin{aligned} RUCC &= \frac{\text{Gross capital expenditure}}{\text{Output}} \\ QRUCC &= \frac{\text{Gross capital expenditure}}{\text{Quality-adjusted output}} \end{aligned}$$

Figure 4.3 shows the trend in $RUCC$ and Figure 4.4 shows the trend in $QRUCC$ since privatisation. The use of the quality-adjusted output measure has only a minor impact on the trend. Capital costs increased in the early years after privatisation and then declined in all sectors up until 1996. The decline was most significant for NGC. After 1996 (ie, the second regulatory period for water and electricity distribution) there was another increase in real unit capital costs. This, again, was followed by a decline towards the end of the regulatory period. We therefore see cyclical variability in this variable, with the cycles affected by the timing of price reviews.

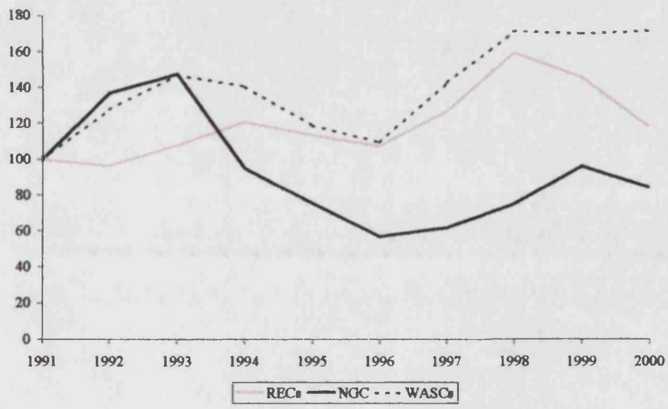

Figure 4.3: RUCC Index

Table 4.2 shows the average annual change in RUCC for each of the sectors. RUCC has increased in all sectors since privatisation, with the largest increase in the water sector. There was a significant decline for NGC in the period 1990 to 1995. When the quality-adjusted volume measure is used, growth is marginally smaller. The growth rate is higher for NGC and the water companies post-1995. It is higher for electricity distribution in the 1990-1995 period. The sectors, particularly water, are characterised by large capital investment programmes which are aimed at improving the quality of supply provided. The improvements may not be delivered for some time however. This is because there is a significant lag between the capital investment and delivery

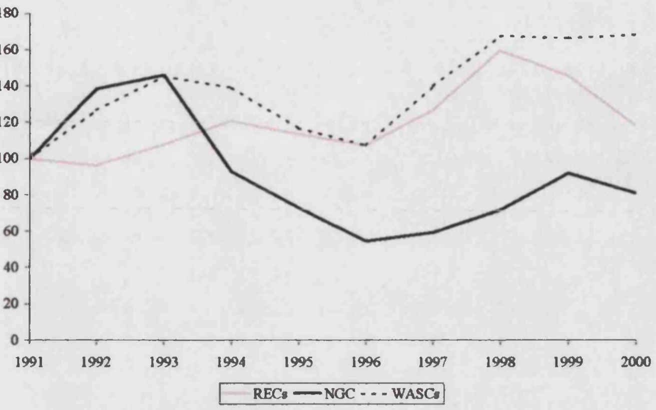

Figure 4.4: QRUCC Index

of supply improvements, with many capital projects taking up to 20 years to complete. This means that there is an increase in capital expenditure today which is not necessarily offset by an increase in outputs delivered.

Table 4.2: Average annual change in RUCC

	1990-95	1995-2000	1991-2000
RUCC			
Distribution	2.51	0.08	1.32
NGC	-7.07	2.25	1.40
Water	2.61	6.35	4.69
QRUCC			
Distribution	2.49	0.07	1.09
NGC	-7.90	2.06	1.06
Water	2.26	6.27	4.49

4.3.3 Analysis

We have found that real unit operating costs have reduced significantly in the water and electricity sectors since privatisation. This indicates that the RPI-X mechanism has delivered on its promise to deliver cost reductions. There has, however, been an increase in real unit capital costs during the period. This reflects the fact that large investment programmes have been undertaken but the quality of supply improvements may not have been delivered yet. In this sense output requirements may be operating against the cost saving incentive. We cannot determine, however, whether the growth in real unit capital costs is higher or lower than it would have been in the absence of price cap regulation. Ian Byatt, the Director General of Water Services, stressed in the 1999 water periodic review that the growth in capital investment was partly financed by efficiency improvements[12]. This suggests that the real unit capital cost growth rate was tempered somewhat by the cost saving incentives.

Our conclusion, therefore, is that real unit operating cost savings have been delivered in the 1990-2000 period and, although real unit capital costs have increased, the extent of the growth may have been reduced by the presence of RPI-X regulation. We examine the particular features of the RPI-X mechanism which affect these trends in unit costs in section 4.6.

The impact of regulation on total costs is difficult to determine from these partial cost measures. The large reduction in real unit operating costs is partially offset by fluctuations in capital costs, as discussed in the literature reviewed in section 4.1, and the overall impact on technical efficiency is unclear. For this reason an analysis of measure which capture the relationship between production and total costs is needed. That is, we wish to capture the impact of all inputs used on outputs rather than relying on single-input measures. This is particularly important when we consider, as discussed in Chapter 3, that the firm may have an incentive to bias its input choices away from non-capital to capital under the RPI-X mechanism. Europe Economics and Crafts (1998) also argue that 'Among different possible ways of measuring productivity, the most relevant for making comparisons between sectors will be TFP, since this is not affected by changes in the relative use of different inputs but only by the overall efficiency with which the inputs are used'. Bearing this in mind, we analyse the relationship between inputs and outputs in more detail using TFP measures.

[12]This is confirmed by Green and Haskel (2001). They note that price increases in the water industry, driven by increased capital expenditure, could indicate the absence of efficiency gains. They note, however, that the regulator believes that 'efficiency improvements actually halved the price increases which would otherwise have been required between 1995 and 2000'.

4.4 Measuring growth in TFP

In this section we explain each of the methodologies used to calculated TFP growth. The growth rate estimates are then presented in section 4.5. We calculate the annual rate of growth in TFP for the water, electricity distribution and electricity transmission companies. In its simplest form TFP is measured as the weighted sum of outputs to inputs:

$$TFP = \frac{\sum_{i=1}^{a} w_i y_i}{\sum_{j=1}^{b} v_j x_j}$$

where y_i is output i, x_j is input j, w_i is the weight on output i, and v_j is the weight on input j.

We use two different approaches to estimate the annual change in this measure. The Tornqvist Index assumes the input and output markets are efficient and determines the weights using input shares. We initially calculate these input shares directly, and then estimate them econometrically from the production function[13]. We also use a Malmquist Index to calculate the change in TFP[14]. This uses a linear programming method - data envelopment analysis - to estimate the *optimal* weights on inputs and outputs.

4.4.1 Tornqvist Indices

We use the standard growth accounting approach to calculate the annual change in the Tornqvist Index. We assume that each regulated firm has a Cobb-Douglas production function:

$$Y = AK^{\alpha}L^{\beta}$$

In the production function, the variable A represents the firm's TFP. It is also referred to as the Hicks neutral shift parameter. K is the capital input into the production process and L is the non-capital input. Y is the value of output produced by the firm. α is the elasticity of output with respect to capital, and β is the elasticity of output with respect to non-capital inputs. These parameters are also the input shares in the Tornqvist Index. We know that if $\alpha + \beta > 1$ there are increasing returns to scale, if $\alpha + \beta < 1$ there are decreasing returns to scale, and if $\alpha + \beta = 1$ there are constant returns to scale. We do not make any a priori assumptions about the returns to scale of the regulated firm's technology.

Taking logs we estimate the firm's TFP as:

[13] It was not possible to estimate a production function for NGC so we only have a Tornqvist Index based on calculated input shares for electricity transmission.

[14] As comparator firms are needed to calculate Malquist Indices, these TFP estimates were only available for the water and electricity distribution sectors.

$$InA = InY - \alpha InK - \beta InL \qquad \text{TFP}$$

The annual growth in TFP is then calculated as[15]:

$$\Delta InA = \Delta InY - \alpha \Delta InK - \beta \Delta InL \qquad \text{TFP Growth}$$

Table 4.3 summarises how each of these variables are measured. There are two different output variables and two different sets of parameters. We therefore have four measures of TFP based on the Tornqvist Index approach.

Table 4.3: Variables used for calculating TFP growth

Variable	Possible measures
Output level (Y)	Volume delivered[1]
	Output adjusted for quality level (year-on-year 'Good quality')
Capital level (K)	Current cost value of gross fixed assets (minus revaluations)
Labour level (L)	Controllable operating costs
Capital costs (rK)	(Treasury rate*Capital level)+Capital charges[2]
Non-capital costs (wL)	Controllable operating costs
Value of output (pY)	Turnover
Parameters	Calculated input shares
	Regression estimate of input shares

Notes to Table: (1) Number of properties for the WASCs. (2) Capital charges are depreciation and infrastructure renewals charges for the WASCs.

The output and input variables are taken directly from the sources discussed in Appendix C. We also need to determine the values of the Cobb-Douglas parameters, α and β. We use two different methods to calculate these parameters, both of which introduce potential biases into the estimates.

1. We calculate the parameters using the share of capital and non-capital costs in turnover.

2. We estimate the production function using regression analysis and use the estimated coefficients for our parameter values.

Each approach is explained below. The calculated TFP growth rates are presented in section 4.5.

[15] $\Delta InA = InA_t - InA_{t-1}$;
$\Delta InY = InY_t - InY_{t-1}$;
$\Delta InK = InK_t - InK_{t-1}$; and
$\Delta InL = InL_t - InL_{t-1}$.

The Tornqvist Index approach may produce biased TFP estimates because of the limiting assumptions which are made to aid the calculations. The main assumptions which may be of concern are as follows[16]:

- production is undertaken using a Cobb-Douglas technology;
- output prices are equal to marginal cost;
- input prices are equal to the value of the marginal product of the input; and
- the only output measure used is the physical number of units produced; no account is taken of changes in the quality of the output provided[17].

We do not know what each firm's technology is, and the Cobb-Douglas restriction is generally regarded as an acceptable assumption in the literature. This restriction means that the actual TFP numbers may not be 100% accurate, but the calculations provide us with a useful guide on changes in productivity over time and relative differences across sectors. The second and third assumptions are only required when the parameters are calculated directly using data on input shares. This approach is adopted below. We also calculate the parameters by econometrically estimating the production function for each sector. In this situation the assumptions on efficient output and input prices are no longer required. Finally, we have attempted to deal with the last problem by adjusting output for the level of quality provided.

Calculated input shares

We initially assume that the firm chooses the level of inputs which maximise profits. We also assume that the product, capital and labour markets are perfectly competitive. Under these conditions a profit-maximising firm chooses the amount of labour at which the wage rate is equal to the marginal revenue product of workers, and the amount of capital at which the cost of capital is equal to the marginal revenue product of capital. That is, as summarised by Hulten (2000), each input is paid its marginal product and hence the output elasticities, α and β, can be substituted by income shares. The

[16]Many of these concerns are emphasised by Weyman-Jones (2001b) who argues that 'The underlying assumptions are unlikely to apply to the analysis of privatised utilities in network industries, both because of their residual market power and because of their known history of productive inefficiency under state ownership'.

[17]This is noted by Hulten (2000) who stresses that 'the TFP residual is intended to measure only the production of more goods - this is what a shift in the production function means - and only the costless portion at that. Innovation that results in better goods is not part of the TFP story'. Waddams Price (2000a) also stresses the need to develop new dimensions into the productivity analysis - such as quality and reliability of service - to ensure that the incentives of the regulatory regime are 'appropriately directed'.

main assumption here is that prices are reasonable proxies for marginal products (ie, products are sold at marginal cost prices). If this is not the case, the TFP estimates calculated using this approach will be biased.

Green and Haskel (2001) support the use of TFP measures for productivity, but they recognise that 'factor shares may not reflect output elasticities if firms have market power, so that if market power changes with privatisation then measured TFP may change for reasons not to do with efficiency'. This raises the question of whether the use of income shares is appropriate for the regulated monopoly firms. In our analysis the monopoly firms have market power, but regulation is expected to control this through the use of price caps ($\widehat{p}$). In particular, for the period which we are analysing, price caps have always been in place. We argue that, as the firm essentially takes its price as given under the regulatory regime, its choices are constrained in a similar way to the price-taker in a competitive market.

Hulten (2000) also discusses the assumption, underlying the TFP calculation, that technological change affects the marginal productivity of all inputs in the same way. This is unlikely to apply in the utilities with factor-specific innovations being more prevalent - eg, engineering changes which improve the productivity of the network assets, or corporate changes which allow for labour-saving devices to be introduced. This means that the TFP measure will be biased. Hulten (2000) suggests using a 'factor-augmentation' production function to deal with this bias. It is not clear, however, how this function would be estimated without ex-ante information on the rates of factor augmentation. We proceed with the standard TFP measure noting that it may be biased.

Bearing these limitations in mind we calculate α and β directly as:

$$\begin{aligned} \alpha &= \frac{rK}{\widehat{p}Y} = \frac{\text{Capital costs}}{\text{Value of output}} \\ \beta &= \frac{wL}{\widehat{p}Y} = \frac{\text{Non-capital costs}}{\text{Value of output}} \end{aligned}$$

Appendix C explains how capital costs, non-capital costs and the value of output are measured. Table 4.4 shows the value of these input shares for the electricity and water sectors. The capital share is higher than the non-capital share in all sectors. The capital share is similar in the electricity distribution and electricity transmission sectors. In the electricity transmission sector the non-capital share is very small. The calculations suggest that there are decreasing returns to scale in the electricity sectors, but constant returns to scale in the water sector.

Table 4.4: Average calculated input shares for the period 1991-2000

	Capital share	Non-capital share	Returns to scale
Distribution	0.462	0.328	0.790
NGC	0.445	0.087	0.532
Water	0.904	0.175	1.079

The TFP growth rates calculated using these input shares are presented in section 4.5.

Estimated input shares

The assumptions regarding marginal cost pricing in the income share approach may not hold in the regulated sectors. We therefore also estimate the input share parameters using regression techniques. The estimated production function coefficients, $\widehat{\alpha}$ and $\widehat{\beta}$, are used as the input shares in the TFP growth equation.

The equation which we estimate is:

$$y_{it} = a + \alpha k_{it} + \beta l_{it} + t + \eta_i + \varepsilon_{it}$$

where y is the log of volume produced, k is the log of capital, l is the log of non-capital inputs, t are time dummies, η is an unobservable firm-specific effect, and ε is the random error term[18]. The coefficients on capital and non-capital are expected to be non-negative.

We estimate separate production functions for the water sector and for the electricity distribution sector. There is an implicit assumption here that each firm within a sector has the same technology and, hence, the estimated sectoral parameters can be used to calculate each firm's TFP growth rate. There is insufficient data to estimate a separate production function for NGC and hence we do not calculate Tornqvist indices on this basis for the electricity transmission sector[19].

The production functions are estimated using the following techniques:

- ordinary least square (OLS);

[18]Baltagi (1995) suggests that the firm-specific effect may represent 'unobservable entrepreneurial or managerial skills of the firm's executives'.

[19]We considered using the distribution production function parameters as a proxy measure for NGC's estimated input shares. It was decided that this was inappropriate as the technologies and cost structures of the two network levels are different and the production functions would therefore be different. Burns and Weyman Jones (1996) argue that the distribution network may be viewed 'as a more complex version of a transmission system since the costs of both types of network are influenced by spatial and geographic factors such as the size of the area served, the density of customers, and so on'. They stress, however, that costs will be different because distribution production is not affected by 'where the power comes from, and how much it costs'. The generation sector has more of an impact on transmission costs. We therefore expect there to be some differences in the underlying production functions.

- OLS on the first-differenced production function (FDOLS);
- within-groups estimation (WG); and
- Anderson-Hsiao estimation (AH).

We describe each estimation technique and its properties in Appendix D. We present estimation results for the standard production function and for the constant returns to scale production function ($\beta = 1-\alpha$). Appendix D also provides an analysis of the variables in the production functions. We find that capital and labour, and the constant returns to scale variable capital/labour, are non-stochastic. There is also a possibility that the error term is correlated over time. This means that the capital and labour regressors will be correlated with the error term in the production function.

With these endogenous explanatory variables, the parameter estimates calculated using OLS, first-differenced OLS and within group estimation are inconsistent. The Anderson-Hsiao estimates are, however, consistent. We therefore use this approach to estimate the parameters of our production function. With this methodology the production function is differenced, thereby removing the firm-specific effect, and instrumental variables are used to estimate the model given the endogeneity of capital and labour in the production function.

We find, in our detailed analysis in Appendix D, that it is reasonable to impose a constant returns to scale restriction on the production functions. Indeed, we only obtain reasonable estimates from our data set when this restriction is used. The first-differenced equation which is estimated is:

$$\Delta(y-l)_{it} = \alpha\Delta(k-l)_{it} + \Delta\varepsilon_{it}$$

We estimate this equation using a number of different instruments for the endogenous regressor. Details of the instruments which were used are presented in Appendix D. We choose the most preferred instrument set for each sector on the grounds that it satisfies the Sargan Test and it provides reasonable coefficient estimates. In both the water and electricity distribution sectors, the preferred instrument set is the level of $(k-l)$ lagged two- and three-periods.

The Anderson-Hsiao estimates for the coefficients in the constant returns to scale production functions are presented in Table 4.5. Standard errors are presented in brackets. In both sectors the estimated coefficients are statistically significant. We also see that the instruments used are valid as the Sargan Test is satisfied[20]. There may be heteroscedasticity in the electricity distribution and water data and, hence, the

[20]The p-value on the Sargan statistic indicates that there is a high probability that the null hypothesis that the instruments are valid is true.

estimates may not be efficient. There is no evidence of autocorrelation in either sector. We note that the estimated electricity distribution capital share is much higher than the capital share calculated using the income share approach. The capital shares in the water sector are similar under both approaches. The TFP growth rates calculated using these estimated input shares are presented in section 4.5.

Table 4.5: Estimated input shares

	Distribution	Water
Capital share	0.921*	0.864*
	(0.880)	(0.135)
Labour share (1-capital share)	0.079	0.136
R^2 (overall)	0.82	0.42
χ^2 statistic	254.6	176.8
Sargan statistic	0.15	0.74
p-value on Sargan statistic	0.70	0.39
No of observations	72	60

Summary

We calculate the growth in TFP as:

$$\Delta TFP = \Delta InY - \alpha \Delta InK - \beta \Delta InL \qquad \text{TFP Growth}$$

Two measures are used for the level of output (Y) - the volume of output delivered and the quality-adjusted level of output. The capital level (K) is measured as the value of gross fixed assets, and the level of non-capital inputs (L) is set equal to controllable operating costs.

We estimate the parameters in this equation, α and β, in two ways. First, the parameters are calculated directly using the ratio of capital and non-capital costs to turnover. This income share approach is only valid if there is marginal cost pricing in the output market and if input markets are perfectly competitive. Given that these conditions do not hold in the regulated utility sectors, we also estimate the production functions directly and use the coefficient estimates as our parameters in the calculation of TFP growth rates.

We note that the first set of parameters are firm-specific but, in the second case, the parameter estimates are sectoral and the same parameter is used for each firm in the sector when estimating its TFP growth rate. In addition, a constant returns to scale restriction is imposed in the econometric analysis but no assumption is made about the returns to scale in the income share calculations.

The calculated parameters are presented above in Tables 4.4 and 4.5. We see that the parameters are quite similar for the water sector. This may be because the assumption of constant returns to scale is reasonable here. In contrast, the estimated parameter in the electricity distribution sector is much higher than the capital share calculated using the income share approach. It is not clear which, if either, of these capital share estimates is most reasonable. The TFP growth rates calculated using these estimated input shares are presented in section 4.5.

4.4.2 Malmquist Indices

We also measure the growth in TFP in the water and electricity distribution sectors using Malmquist Indices. These Indices are calculated using data envelopment analysis and no assumptions are made about the firm's technology, or competition in the product and input markets. We explain here how the indices are calculated and the corresponding TFP growth rates are presented in section 4.5. We are unable to calculate TFP growth with this methodology for NGC as there is only one transmission company in England and Wales and no suitable comparators on which to base the estimation.

The idea behind Malmquist Indices originates from Farrell (1957). He used distance-measures to determine the economic efficiency of a firm. The measures, which incorporate technical efficiency and allocative efficiency, are illustrated in Figure 4.5 for a two output and single input technology[21]. The production possibility frontier (PPF) indicates the maximum amount of output combinations, (y_1, y_2), which can be produced with the given input (x). The isorevenue line shows the amount of revenue which can be earned from different output combinations given existing prices. The firm operating at point A is technically inefficient (below the PPF) and allocatively inefficient (below the isorevenue line). Technical efficiency is measured by the distance ratio $\frac{0A}{0B}$ and allocative efficiency is measured by the distance ratio $\frac{0B}{0C}$. Overall economic efficiency is then measured as the product of these measures giving us:

$$\text{Efficiency} = \frac{0A}{0C}$$

The position of the PPF can be estimated using Data Envelopment Analysis (DEA). More specifically, we use linear programming techniques to construct a piece-wise frontier or 'envelope' so that all data points in the sample lie on or below the frontier. The frontier is constructed so that the input and output weights in the TFP formula are maximised subject to the constraints that the efficiency measure must be

[21] Technical efficiency measures the firm's ability to maximise the level of output produced for a given set of inputs. Allocative efficiency measures the extent to which the firm uses its inputs in optimal proportions given the current technology and input prices.

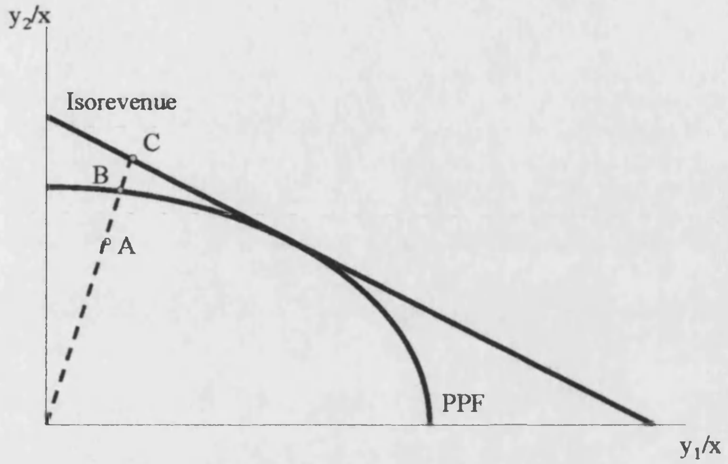

Figure 4.5: Measuring efficiency

less than or equal to one for each data point. This contrasts with the Tornqvist Index where calculated weights, based on observed data, are used.

When panel data is available, Malmquist Indices can be constructed using these DEA frontiers. The indices provide a measure of annual TFP change for each firm and each year in the panel. The Malmquist Index calculates distance functions relative to the current PPF and relative to the PPF in the previous year. The relevant formula, for TFP change between period s and period t, is a geometric mean of the current and previous period TFP indices[22],[23]:

$$m = \left[\frac{d^s(y_t, x_t)}{d^s(y_s, x_s)} \times \frac{d^t(y_t, x_t)}{d^t(y_s, x_s)}\right]^{1/2}$$

where y is the set of outputs, x is the set of inputs and $d^j(y_i, x_i)$ is the output-orientated distance function given the period j technology and period i inputs and outputs. If m is greater than 1 there has been growth in TFP. If it is less than one there has been a decline in TFP.

Figure 4.6 illustrates the distance functions included in the Malmquist Index for a constant returns to scale technology involving a single output and a single input. The firm is at point s in period s and point t in period t. In both periods the firm is inefficient - ie, below the frontier. There has been an advance in technology between period s and period t as illustrated by the change in the position of the frontier.

Each distance function is calculated using linear programming methods. Details

[22]Fare et al (1994) proposed this geometric index. They note that 'this form is typical of Fisher ideal indexes'.

[23]The notation used here is borrowed from Coelli et al (1998).

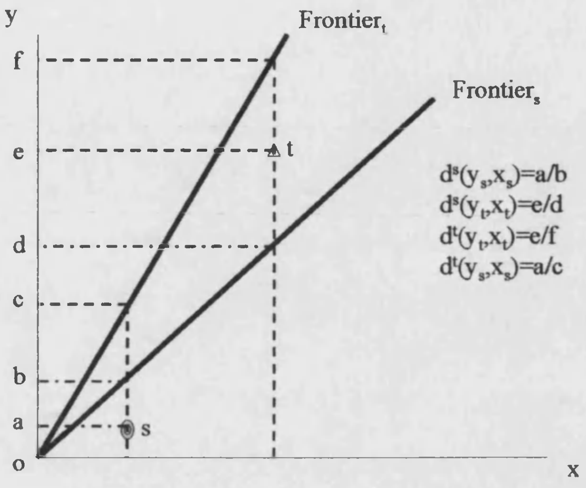

Figure 4.6: Distance functions

of the process can be found in Coelli et al (1998). The linear program is solved four times for each firm - for period i inputs relative to period j's technology set; for period i inputs relative to period i's technology set; for period j inputs relative to period i's technology set; and for period j inputs relative to period j's technology set. The four programmes are run for each firm for each pair of adjacent time periods. Constant returns to scale is assumed in the calculation of the distance functions using DEA.

The main advantages of Malmquist Indices over Tornqvist Indices are as follows[24].

- Cost or price data are not required. This means that income shares do not need to be estimated and hence we do not run into problems about how to calculate these shares.
- There is no assumption that firms are profit-maximisers or cost-minimisers.
- We do not impose any assumption (eg, Cobb-Douglas) on the functional form of the production function[25].
- The entire panel data set can be used to calculate TFP change, and the change can be decomposed into efficiency change (ie, firms catching up with the efficiency

[24]The discussion of the advantages and disadvantages of Malmquist Indices is largely taken from Coelli et al (1998). Weyman-Jones (2001a) and Jamasb and Pollitt (2000) provide similar arguments for and against Data Envelopment Analysis. A detailed discussion of this approach can also be found in Majumdar (1997).

[25]Fare et al (1994) show that with a Cobb-Douglas production function the Malmquist Index is equal to the 'ratio of the efficiency parameters' of the function.

frontier) and technical change (ie, shifts in the efficiency frontier)[26].

The main disadvantages of this approach, many of which also arise with econometric estimation of the production function, are:

- The shape and/or position of the DEA frontier can be affected by measurement error.
- The results will be biased if any inputs or outputs are excluded.
- The efficiency scores are determined relative to the best performing companies in the current sample. Introducing extra firms can therefore reduce the score. In addition, efficiency scores can't be compared across studies which have different samples.
- The linear programming method is 'data-hungry'. That is, it operates best with large datasets. If there are only a few firms, and many inputs and outputs, most of the firms will appear on the frontier.

We measure TFP growth in the regulated water and electricity distribution sectors using these Malmquist Indices[27]. We construct two indices; each of which has one output and two inputs. The inputs used are the capital level and the non-capital level. These measures are discussed in Appendix C. We use two different output measures - volume and quality-adjusted volume.

4.5 TFP Growth - the evidence

We use the three methodologies discussed in section 4.4 to calculate the rate of growth in TFP in the regulated sectors. The average annual growth rates for the electricity distribution and water sectors are presented in Tables 4.6 and 4.8. We provide data on

[26]The Malmquist Index formula can be decomposed into two parts:

$$\begin{aligned} m &= \left[\frac{d^s(y_t, x_t)}{d^s(y_s, x_s)} \times \frac{d^t(y_t, x_t)}{d^t(y_s, x_s)}\right]^{1/2} \\ &= \frac{d^t(y_t, x_t)}{d^s(y_s, x_s)} \left[\frac{d^s(y_t, x_t)}{d^t(y_t, x_t)} \times \frac{d^s(y_s, x_s)}{d^t(y_s, x_s)}\right]^{1/2} \end{aligned}$$

where $\frac{d^t(y_t,x_t)}{d^s(y_s,x_s)}$ is a measure of how the firm's distance function has changed relative to the current PPF from one period to the next (efficiency change) and $\frac{d^s(y_t,x_t)}{d^t(y_t,x_t)} \times \frac{d^s(y_s,x_s)}{d^t(y_s,x_s)}$ is a measure of how the PPF has changed from one period to the next (technical change).

[27]The Malmquist Indices were constructed using ©DEAP2.1. This programme was kindly provided by Tim Coelli - www.une.edu.au/econometrics/cepa.htm. Details about this programme can be found in Coelli (1996).

the average growth rate for the entire period since privatisation (1991-2000), and for each of the regulatory periods 1991-1995 and 1995-2000. NGC's average annual TFP growth rates are presented in Table 4.7. Again, the data is provided for the entire period since privatisation (1991-2000) and for each of NGC's regulatory periods; 1991-1993, 1993-1997 and 1997-2000. The sectoral trends in TFP growth are shown in Figures 4.7 to 4.9 at the end of this section. These charts are based on the standard output measure but we note that the trends over time do not change significantly when the quality-adjusted output measure is used. We summarise the main findings from each approach here.

Table 4.6: Average annual TFP growth rates in the electricity distribution sector

	91-00	91-95	95-00
Tornqvist Indices (%)			
Calculated shares			
Standard output measure	1.97	1.62	2.25
Quality adjusted output measure	1.98	1.48	2.38
Estimated shares			
Standard output measure	0.88	0.84	0.92
Quality adjusted output measure	0.90	0.89	0.82
Malmquist Indices (%)			
Standard output measure	2.50	2.03	2.88
Quality adjusted output measure	4.13	3.30	4.80

Table 4.7: Average annual TFP growth rates for NGC

	91-00	91-93	93-97	97-00
Tornqvist Indices (%)				
Calculated shares				
Standard output measure	1.67	-1.77	3.28	1.82
Quality adjusted output measure	2.14	-1.23	4.02	1.88

Table 4.8: Average annual TFP growth rates in the water sector

	91-00	91-95	95-00
Tornqvist Indices (%)			
Calculated shares			
Standard output measure	0.95	2.08	0.04
Quality adjusted output measure	1.15	2.44	0.12
Estimated shares			
Standard output measure	0.92	2.07	0.00
Quality adjusted output measure	1.12	2.43	0.08
Malmquist Indices (%)			
Standard output measure	3.68	8.23	0.04
Quality adjusted output measure	5.40	12.45	-0.24

Tornqvist Indices based on calculated shares

The average annual change in TFP is positive but reasonably small in all sectors when calculated using the unadjusted output variable. The average growth rate since privatisation is highest in the electricity distribution sector and lowest in the water sector.

The electricity distribution sector average growth rate was higher in the second regulatory period (1995-2000) than the first. These averages hide the fact that there was a steady increase in the growth rate up to 1996, and that this was followed by a decline in the rate up to 1998. There was a slight improvement in the rate of TFP growth after this, but the size of the improvement was well below the rates experienced in the first regulatory period.

In contrast, the average growth rate was higher in the water sector immediately after privatisation (1991-1995). This is largely due to the high growth rate in 1993 which was driven by the large drop in labour costs at the time. The growth rates have fluctuated significantly over time, with an increase in 1995 and a notable reduction in productivity (negative TFP growth rates) between 1997 and 1999. These fluctuations coincide with the periods just after the end of a price review and just before a price review, respectively.

NGC's average annual TFP growth rate has also fluctuated over time. There was negative growth in 1992 but this was followed by positive, and increasing growth in all other years. There was an average decline in TFP in the first regulatory period (1991-1993). This was followed by large TFP growth rates in the second period (1993-1997). Productivity growth continued in the third period (1997-2000), but the rate of change was lower than in the second period.

When the quality-adjusted output measure is used, the trends in TFP growth remain the same. In general the size of the growth rates are higher, confirming that the firm is delivering 'more' with the given level of inputs. The difference is most significant for NGC, and for the water sector between 1991 and 1995. The only case where we get lower TFP growth rates with the adjusted output measure is for the electricity distribution sector between 1991 and 1995.

Tornqvist Indices based on estimated shares

We find positive average annual TFP growth in all sectors with this methodology. The growth rates are lower than those based on the calculated shares, significantly so for electricity distribution. The trend in the growth rate is the same for both sectors as it was for the calculated share measure, as seen in figure 4.8.

We again find that the use of the quality-adjusted output measure increases the level of TFP growth slightly in all sectors, but it does not change the general trends in the growth rates. The only exception is the electricity sector between 1995 and 2000, where the growth rate is again smaller with the adjusted output measure than with the unadjusted output measure.

Malmquist Indices

Our calculations based on this methodology also indicate that there has been an average annual improvement in TFP in all the sectors since privatisation. The growth rates are significantly larger than those calculated using the Tornqvist Indices in both the electricity distribution and water sectors. The water sector has the highest rate of improvement. These growth rates may be biased by the small sample size for the DEA calculations.

The trends in TFP growth rates are similar to those from the Tornqvist measures. Productivity has fluctuated in the electricity distribution sector, with a steady increase up to 1995, a step-jump improvement in 1996, followed by a decline up to 1998 and an improvement in 1999. The pattern of change has also been stark in the water sector. There was significant TFP growth between 1990 and 1995, primarily driven by the large decline in labour costs in 1993. This was followed by a period of very low, almost zero, growth between 1995 and 2000. This low level of growth was driven by negative or zero TFP growth rates in 1998, 1999 and 2000.

The improvement in productivity is significantly higher in both the water and electricity distribution sectors when the quality-adjusted output measure is used. This is a marked difference from the Tornqvist Index measures. The trends over time are not affected by the use of the quality-adjusted output measure however.

Summary

We conclude that there has been some improvement in TFP in each of the regulated sectors since privatisation.

- Average annual TFP growth in the electricity distribution sector ranged from 0.88% to 2.50% for the period 1991 to 2000, depending on which measure was used to calculate TFP. The annual reduction in real unit operating costs for the same period was comparable at 1.32%. The range was wider - from 0.9% to 4.13% - when quality-adjusted output was used. The reduction in real unit operating costs based on the quality-adjusted output measure also lay within this range at 1.09%. For nearly all measures TFP growth was higher in the period 1995 to 2000 than for the first regulatory period. This is consistent with the literature discussed in section 4.1.

- NGC's annual TFP growth rate was 1.67% for the period 1991-2000, which is slightly higher than the annual reduction in real unit operating costs of 1.4%. When quality-adjusted output measures are used, average annual TFP growth was 2.14% and the reduction in real unit operating costs was again lower at 1.06%. NGC's productivity improvement was highest during the second regulatory period (1993 to 1997).

- In the water sector, average annual TFP growth for the period 1991 to 2000 was in the range 0.92% to 3.68%. The reduction in real unit operating costs was higher at 4.69%, suggesting that the reduction in operating costs is at least partially offset by increased capital costs. This is consistent with the literature discussed in section 4.1. The range was higher, but of a similar width, when the quality-adjusted output measure was used: 1.12% to 5.4%. The quality-adjusted real unit operating cost reduction was just inside this range at 4.49% per annum. Productivity improvement was higher for the period 1991 to 1995 than for 1995 to 2000.

In general, the average annual growth rates have been reasonably small. The rates are slightly higher when the quality-adjusted output measure is used. In the water sector the rates of productivity improvement contrast significantly with the large reductions in real unit operating costs but in the electricity sectors the annual efficiency improvements reflected in TFP growth rates and in real unit operating costs reductions have been comparable. The rates of TFP growth vary over time in each of the sectors, and in some cases appear to have been affected by the timing of regulatory reviews. For example, there is evidence of productivity improvements immediately after a periodic review (e.g. in 1995 for water) and for productivity to be slowed down just before a periodic review.

The trends in the growth rates do not vary significantly when we use different methodologies to estimate the TFP. However the growth rates are significantly higher when the Malmquist methodology is used instead of the Tornqvist Indices. The growth levels are reasonably similar from the calculated Tornqvist Indices and the estimated Tornqvist Indices however. This indicates that the choice of methodology will have an impact on conclusions reached about the rate of productivity improvement in the regulated sectors since privatisation. These differences are only a matter of degree in the level of growth, however, and we stress that the general trend in growth rates does not change significantly from one methodology to another. This can be seen in figures 4.7 to 4.9.

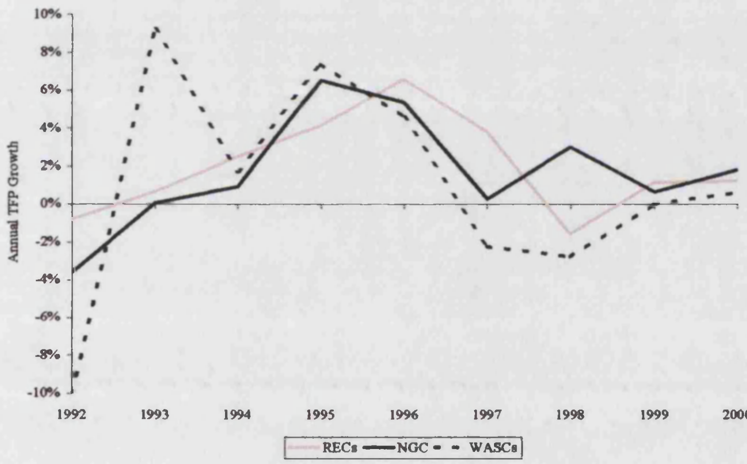

Figure 4.7: TFP Growth: Tornqvist Indices using calculated shares

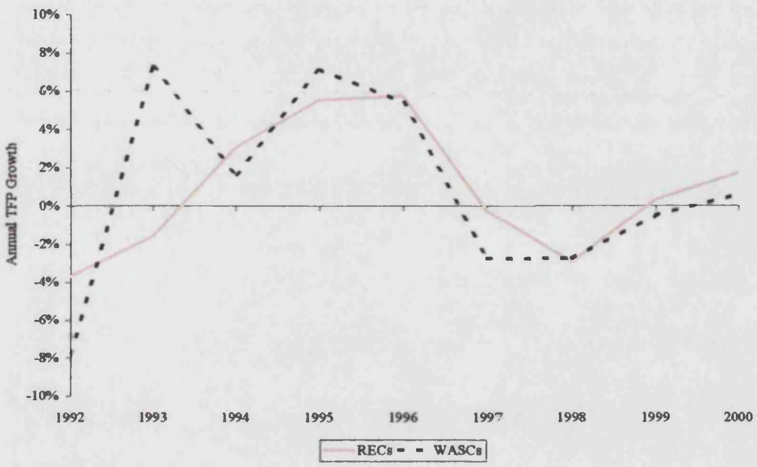

Figure 4.8: TFP Growth: Tornqvist Indices using estimated shares

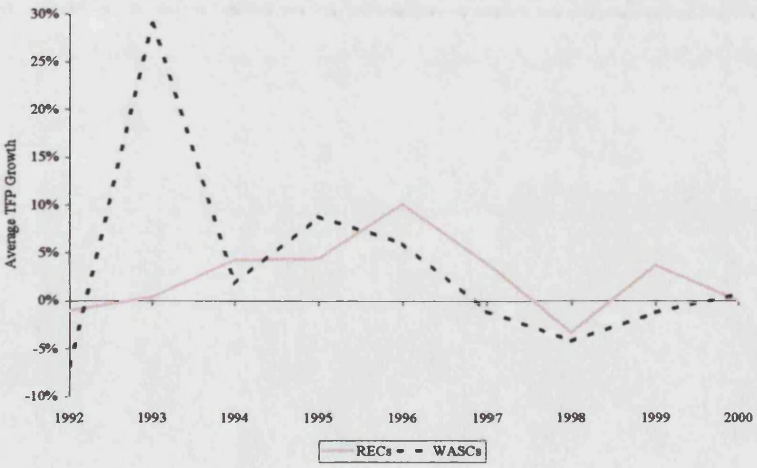

Figure 4.9: TFP Growth: Malmquist Indices

4.6 Impact of RPI-X Regulation

There has been a reduction in real unit operating costs, and an increase in total factor productivity, across the water and electricity sectors since privatisation. We wish to determine how, if at all, these productivity changes have been affected by RPI-X regulation. The main aspects of the RPI-X regime which are expected to impact on the change in productivity are outlined here[28]. Many of the points made are discussed in more detail in Chapters 2 and 3.

- The price cap mechanism provides the firm with an incentive to reduce costs. The higher the share of efficiency savings which a firm expects the retain, the greater the incentive to reduce costs. We found, in section 3.1.2, that this incentive is dulled by two key features of the price-cap setting process. First, the firm's decision about how much cost reduction to undertake is determined by the length of time over which it expects to be allowed to retain the profits. This length of time is called the regulatory lag. Second, the regulator will intervene to change the regulatory contract mid-period with positive probability. This is often done when the regulator wishes to share the benefits of cost savings with consumers early. The absence of commitment to the contract reduces the length of time that profits are retained for and, hence, reduces the net benefit from making cost savings in the first place. The incentive to reduce costs is determined by the interaction of these two factors, as the expected regulatory lag will be affected by the extent of regulatory commitment. We therefore include the *joint impact* of these two variables as a measure of the expected share of efficiency savings.

- Since privatisation the regulated firms have been under increasing pressure to deliver a higher quality of service to consumers. Any attempts to reduce costs may therefore have been offset, at least partially, by the demand for more quality. Of course, as discussed in section 3.1.3, the incentive to reduce costs may have worked in the other direction, with the firm choosing to underdeliver on quality.

  The overall impact of the output targets on productivity is ambiguous. Productivity improvement would arise if the firm adjusted its production processes so that the higher level of output was delivered with the same or lower levels of inputs. There would also be evidence of productivity improvement, however, if

[28]We note, as an aside, that the level of the price cap does not impact on the incentive to improve productivity. So long as the price cap is not so severe as to endanger the financial viability of the firm, there will always be an incentive, no matter what the value of X, to increase efficiency levels beyond those assumed by the regulator. The incentive is determined by 'the right of the company to any residual profit' (Markou and Waddams Price, 1999). It is therefore the methodology used to share savings at periodic reviews, rather than the size of the X-factor, which is expected to affect the rate of change in productivity.

both the level of output and the level of inputs were reduced. This would be at the expense of quality of service. We therefore need to examine the impact of the quality of service level on trends in productivity.

- The incentive to make operating cost savings does not necessarily translate into an incentive to become more productive. As noted in section 4.3, capital cost increases may offset a reduction in operating costs, some forms of operating costs (eg, the costs of contracting out work) may increase while others (eg, direct labour costs) may decrease, and cost reduction in one year may be offset by increased costs the next year[29]. In addition, as noted above, the incentive to reduce costs may be offset by an increase in required outputs.

  For these reasons, the regulatory regime, which is primarily focused on delivering operating cost savings, may not necessarily improve the firm's overall productivity level. We therefore examine the impact of the regulatory regime on changes in real unit operating costs and on TFP growth rates. We expect to find, given our analysis of the RPI-X regime and evidence in the literature discussed in section 4.1, that the regime has a greater impact on unit operating costs than on TFP.

The rate of productivity growth is also affected by factors outside of the RPI-X regime. We focus on two specific factors here. First, the firm's productivity is expected to improve after privatisation because the change of ownership provides new cost reduction incentives[30]. We wish to determine how privatisation affects the change in productivity over time. In particular, we wish to see whether the impact reduces after a number of years. The second factor of interest also relates to the firm's ownership. We consider the impact that takeovers or mergers may have on the firm's productivity performance. It is assumed that a primary reason for undertaking such corporate activity is to reduce costs and hence we expect that the corporate activity will improve future productivity.

We have identified a number of factors within the RPI-X regime, and two factors relating to ownership, which may have had an impact on productivity growth trends in the water and electricity sectors since privatisation. We test the relationship between

[29] Markou and Waddams Price (1999) present a formal model which shows how the RPI-X mechanism will 'lead to a significant increase in capital and decrease in labour employed after privatisation, and more particularly a sharp rise in labour productivity'.

[30] Vickers and Yarrow (1988a) provide a detailed discussion of the expected theoretical implications of a transfer from public to private ownership. The reader is also referred to Green and Haskel (2001), Helm and Yarrow (1988), Markou and Waddams Price (1999), Pollitt (1999), Yarrow and Jasiński (1996) and Yarrow (1989) for discussions of the expected advantages and disadvantages of privatisation.

these variables by estimating the following equation using annual data on each of the regulated firms:

$$\Delta prod_{it} = \beta_1(effshare) + \beta_2 priv_{it} + \beta_3 corp_{it} + \beta_4 qual_{it} + t + \varepsilon_{it}$$

where:

- $\Delta prod$ is the annual change in productivity, measured here as the growth in TFP and as the reduction in real unit operating costs (RUOC).
- *effshare* captures the firm's expectation of the share of efficiency savings which it will retain. It is equal to the product of the expected regulatory lag and the degree of regulatory commitment ($lag \times comm$).

  lag is the firm's expectation of the length of the regulatory lag. We expect that the longer the regulatory lag, the greater the level of reduction in real unit operating costs.

  comm is a commitment index which represents the firm's belief about how committed the regulator is to the current regulatory contract. The more committed the regulator is to the regulatory contract, the higher is the firm's expectation of the profit it earns from each unit of effort, and hence the higher the level of cost reduction.

  We expect the combined term to have a positive impact on productivity growth. TFP will increase with an increase in the expected efficiency share and the change in real unit operating costs will decline - get more negative - with an increase in the expected efficiency share.

- *priv* is the number of years since privatisation. If privatisation is the main factor affecting cost savings and productivity, we expect the rate of improvement (an increase in TFP or a decrease in real unit operating costs) to decline the longer the time period since privatisation. In contrast, if regulation has an ongoing and significant impact, cost savings and productivity improvement would continue to be strong as the private firm matures. The presumption is that productivity growth later in the period would be attributed to the impact of regulation rather than privatisation.
- *corp* is a dummy variable indicating whether or not the regulated firm was involved in a takeover or merger in the previous year. A firm's cost saving potential, and productivity growth rate, are expected to improve if the firm is taken-over and/or if it merges with another company. That is, we expect a positive relationship between TFP growth and this variable and a negative relationship between the real unit operating cost change and the corporate restructuring variable.

- *qual* is a scale variable which represents the level of quality-adjusted output delivered by the firm. The higher this variable the more 'output' the firm is delivering with its inputs. This may constrain productivity improvements.

- t is a time dummy variable.

We explain in Appendix E how each of these explanatory variables are constructed. The estimation results are presented here.

4.6.1 Estimation results

We present below the results of our estimation. The data used is a pooled cross-section for the electricity distribution, electricity transmission and water sectors. We have twenty-three companies and data over nine years, giving us 207 observations in total[31],[32]. There are numerous problems with our results which we believe are driven by the relatively short dataset, a limited amount of variability in the explanatory variables, and the range of missing variables which also affect productivity growth. The low goodness-of-fit measures which arise in most of the models suggest, in themselves, that regulation is not the primary factor affecting productivity change in these sectors. In addition, the extremely high goodness-of fit measures which arise in the Malmquist Index regressions are considered unreliable. There is also likely to be measurement error with both the dependent variables and the explanatory variables. The analysis is therefore considered preliminary at this stage and would warrant further investigation when a longer and more variable dataset is available.

We estimate the equation using an ordinary least squares regression (with robust standard errors)[33]. We present the estimation results for eight different dependent variables - the unadjusted and quality-adjusted change in real unit operating costs, and the three unadjusted and quality-adjusted TFP growth rates discussed in section 4.4.

[31]We have not estimated Tornqvist Indices using estimated shares or Malmquist Indices for NGC. When these variables are used as the dependent variable we have twenty-two companies and nine years of data, giving us 198 observations.

[32]We also estimated the equation using an alternative data set which had the 1992/93 data for all water companies removed. This was done to check whether the large efficiency improvement in this year, arising from the reduction in labour costs, affected the impact regression results significantly. We found that for most measures the goodness-of-fit measures did not change significantly with this alternative data set and for most variables the size and sign of the coefficient estimates did not change by very much. In addition most estimates remained statistically insignificant. The main variation was with the expected efficiency share variable whose coefficient became negative for some Tornqvist measures, which is inconsistent with our expectations. The value of the coefficient continued to be very low, however, and was statistically insignificant.

[33]We also considered estimating the equation using the within group methodology. It was decided that this was inappropriate as there is very little varation in the variables across firms. Most of the variation is over time and/or across sectors.

The results are presented in Table 4.9. Standard errors for each of the coefficient estimates are provided in brackets. We also present the 95% confidence interval for each coefficient in square brackets below the coefficient estimate. A * on a coefficient, or on the F-statistic, indicates that the estimate is statistically significant at the 1% level (ie, it has a p-value which is less than 0.01). We consider these results by discussing each of the productivity measures separately.

Table 4.9: Impact of Regulation: OLS estimates

	Productivitymeasure			
	Tornqvist		Malmquist	RUOC
	Calculated	Estimated		
Standardoutput				
Effshare(lag*commitment)	.0008	.0010	-.0001	-.0237*
	(.0016)	(.0016)	(.0050)	(.0051)
	[-.002,.004]	[-.002,.004]	[-.01,.01]	[-.034,-.014]
Yrs since privatisation	-.0297	-.0227	.2986*	.1105*
	(.0119)	(.0106)	(.0831)	(.0342)
	[-.053,-.006]	[-.043,-.002]	[.136,.461]	[.043,.178]
Corporate activity	.0058	.0092	.0062	.0096
	(.0190)	(.0091)	(.0237)	(.0695)
	[-.031,.043]	[-.009,.027]	[-.04,.053]	[-.127,.146]
Quality-adjustedoutputlevel	.0001	-.0001	.0030	.0001
	(.0002)	(.0001)	(.0017)	(.0005)
	[0,0]	[0,0]	[0,.006]	[-.001,.001]
n	207	198	198	207
R^2	0.37	0.54	0.99	0.32
F-statistic	16.68*	33.25*	9640.91*	22.09*
Quality-adjustedoutput				
Effshare(lag*commitment)	.0011	.0014	.0072	-.0240*
	(.0016)	(.0017)	(.0055)	(.0050)
	[-.002,.004]	[-.002,.005]	[-.004,.018]	[-.034,-.014]
Yrs since pr ivatisation	-.0284	-.0219	.2592	.1093*
	(.0119)	(.0109)	(.0938)	(.0340)
	[-.052,-.005]	[-.043,-.001]	[.075,.443]	[.043,.176]
Corporateactivity	.0046	.0080	.0049	.0108
	(.0189)	(.0109)	(.0410)	(.0694)
	[-.032,.042]	[-.009,.027]	[-.075,.085]	[-.125,.147]
Quality-adjustedoutputlevel	.0000	-.0002	.0034	.0001
	(.0002)	(.0002)	(.0019)	(.0005)
	[0,0]	[-.0001,0]	[0,.007]	[-.001,.001]
n	207	198	198	207
R^2	0.38	0.53	0.98	0.33
F-statistic	16.26*	38.73*	2911.72*	22.23*

Tornqvist Indices with calculated input shares

When this productivity measure is used, all of the coefficient estimates are statistically insignificant, and the goodness-of-fit of the model is quite low. The F-statistic is statistically significant, however, suggesting that the overall model is reasonably appropriate. This is true for both the standard output measure and the quality-adjusted output measure. We find that the coefficient on the efficiency share variable is positive but quite low. The sign of the relationship corresponds with our expectation that the firm will be more productive the longer is the lag and the more commitment there is to the contract but the size of the coefficient suggests that the relationship is not very strong.

The number of years since privatisation has a negative coefficient, suggesting that as the privatisation effect reduces over time it is not compensated by a regulation effect. There is a positive coefficient on the corporate activity variable, as expected, but the size of the correlation is small. There is also a small but positive coefficient on the level of quality-adjusted output delivered.

The results are very similar when the quality-adjusted output TFP measure is used as the dependent variable. The sign of the coefficients are the same for all of the variables, and the size of the coefficients are in general marginally higher with this measure (ie, they become more negative or more positive).

Tornqvist Indices with estimated input shares

With this productivity measure all of the coefficients are statistically insignificant. The goodness-of-fit of the model, as measured by R^2 and the significance of the F-statistic, is reasonable however. There are less observations with this estimation compared to the calculated Tornqvist Indices because of the absence of NGC from the dataset.

The signs of most of the coefficients are the same for both the Tornqvist measures. The only difference is with the quality-adjusted output level, which is negative but still very low in this model. The positive coefficients on the expected efficiency share and corporate activity variables are slightly higher with the estimated input share indices. The coefficient on the privatisation term is less negative.

The results do not change significantly when the quality-adjusted TFP growth rate is used as the dependent variable. The sign of the coefficients are not changed and, in most cases, the size of the coefficient is increased slightly (ie, becomes more negative or more positive). We also note that the comparison between the estimated coefficients in the two Tornqvist models does not change when we consider the quality-adjusted measures.

Malmquist Indices

We get the most promising results when this productivity measure is used. The R^2 value is very high, and some of the coefficient estimates are statistically significant. The goodness-of-fit measure is so high that it is considered an unreliable indicator of the appropriateness of the model.

Surprisingly we find a low negative coefficient on the expected efficiency share variable in the standard model. This contrasts with our expectation that an increase in expected efficiency share would lead to an increase in productivity. The number of years since privatisation has a positive and statistically significant coefficient. This may suggest that regulation has an ongoing impact on productivity growth when the privatisation effect dwindles away. We also find a small but positive coefficient on both corporate activity and the quality-adjusted output level. This is consistent with our expectations.

As with the other TFP measures, the analysis does not change significantly when quality-adjusted output is used. The main exception is that now the efficiency share variable has the expected positive sign, although the size of the coefficient is still relatively small. It is however higher than was the case in the estimation using Tornqvist Indices. The signs of the coefficients on the other variables do not change. The privatisation coefficient is lower than in the standard model and is now statistically insignificant. The corporate activity coefficient is lower, and the coefficient on the quality-adjusted output variable is slightly higher.

Real unit operating costs

Two of the estimated coefficients are statistically significant when the change in real unit operating costs is used as the dependent variable in our estimation. The R^2 value is comparable with that obtained when the Tornqvist Indices are used.

We find, as expected, a negative relationship between changes in real unit operating costs and the expected efficiency share. This indicates that an increase in the efficiency share leads to a decline in real unit operating costs (i.e. a productivity improvement). This estimate is statistically significant and the size of the coefficient is higher than in all other models, suggesting that the regulatory regime has a greater impact on cost reduction than on technical efficiency.

The number of years since privatisation has a positive and statistically significant coefficient, and the size of the coefficient is quite high. This implies that the longer the time since privatisation, the lower the rate of decline in real unit operating costs. This suggests that regulation does not have an ongoing effect on the change in real unit operating costs after the privatisation effect has been reduced. We also find a small positive coefficient on corporate activity which suggests that real unit operating costs increase after corporate restructuring (i.e. there is not an improvement in productivity). Similarly, an increase in the quality-adjusted output level would lead to a small increase in real unit operating costs according to the small but positive estimated coefficient.

When the quality-adjusted output level is used to calculate real unit operating costs the signs of the coefficients do not change. The coefficients are also only marginally different in size with this adjusted measure. Finally, we note that for the efficiency share and privatisation variables the direction of the relationship between the coefficients and the productivity growth measure are the same as when the Tornqvist Indices are used. The relationship is stronger with real unit operating costs however.

4.6.2 Summary

We conclude that the regulatory variables have some impact on productivity growth, but the goodness-of-fit measures suggest that other factors also affect the growth rate. In general the methodology used to calculate the dependent variables affects our analysis and it is difficult to determine an exact relationship between productivity growth and each of the explanatory variables. In addition it is difficult to draw conclusions when most of the coefficient estimates are statistically insignificant.

However, the results can be used as a benchmark guide to the relationship between productivity growth and RPI-X regulation[34]. We summarise the key conclusions here.

- As expected, TFP growth increases with the expected efficiency share. The size of the impact is quite small however - the confidence intervals for the Tornqvist TFP measures suggest that with 95% probability the interval -0.002 to 0.004 contains the true coefficient. The interval with real unit operating costs is slightly higher, -0.034 to -0.014, but the coefficient is still quite low. The may support the argument that regulation has more of an impact on operating cost levels than on technical efficiency.

- For all our productivity measures, except the Malmquist Indices, there is a negative relationship between productivity growth and the number of years since privatisation. This suggests that regulation may not have an ongoing impact on productivity when the privatisation effect dwindles away. The confidence intervals for the Tornqvist TFP measures suggest that with 95% probability the

[34]The reader is reminded that productivity growth is measured as an increase in TFP and a decrease in real unit operating costs. Opposite signs on the coefficients are therefore an indication of a consistent relationship between productivity growth and the explanatory variables when the underlying dependent variable changes.

interval -0.043 to -0.002 contains the true coefficient. The interval with real unit operating costs, 0.043 to 0.178, suggests a higher coefficient. This suggests that regulation may have greater impact on real unit operating costs than on TFP.

- For all our TFP variables, corporate activity is positively correlated with productivity growth but the size of the impact is quite small. There is a negative relationship between productivity growth when the decline in real unit operating costs is used as the dependent variable. The confidence intervals for the Tornqvist TFP measures suggest that with 95% probability the interval -0.031 to -0.043 contains the true coefficient. The interval is higher but quite wide with real unit operating costs; -0.127 to 0.146.

- The higher the level of quality-adjusted output produced, the lower is the level of productivity growth. This suggests that the additional output requirements constrain productivity growth somewhat. The opposite relationship is observed when the Tornqvist Indices with estimated factor shares are used as the dependent variables. In all cases the size of the impact is *very small and potentially non-existent.* The confidence intervals for the Tornqvist TFP measures suggest that with 95% probability a very small interval around zero contains the true coefficient. The interval with real unit operating costs, -0.001 to 0.001, also indicates that the coefficient is low.

- When we use the quality-adjusted productivity measures the results are not altered significantly. In most cases the coefficient signs are the same as when the unadjusted productivity measure is used, but the size of the coefficients are changed marginally. The confidence intervals for all coefficients are very close to their counterparts in the standard models.

In general, therefore, the estimation results are consistent with our analysis in Chapter 3. We reiterate, however, that there are limitations to the model, and the size and direction of the coefficients are partially dependent on the methodology used to calculate TFP. It is therefore difficult to reach any firm conclusions here about the relationship between RPI-X regulation and productivity improvement.

4.7 Conclusion

We examined the extent to which the RPI-X regime has delivered on its promise of improved productivity by calculating costs changes in the water and electricity sectors since privatisation. There has been a significant decline in real unit operating costs in all sectors, but real unit capital costs have increased. The rate of increase may be lower than would be the case in the absence of the incentive regime.

Cost savings do not necessarily imply an improvement in productivity however. We therefore also measured the rate of growth in TFP to see whether the RPI-X regime has succeeded in improving technical efficiency in the regulated water and electricity sectors. Three different methodologies were used to calculate TFP - Tornqvist Indices using input shares which were calculated directly from the firm's income shares; Tornqvist Indices using input shares taken from the estimated sectoral production functions; and Malmquist Indices calculated using Data Envelopment Analysis[35]. We found that, for all methodologies, the rate of growth in TFP has been positive but relatively small, particularly when compared to the reductions in real unit operating costs. We also found that the rate of TFP growth varies significantly over time and it appears to be affected by the timing of periodic reviews.

Our analysis of the trends in productivity growth does not change significantly when the growth rates are calculated using quality-adjusted output levels. The actual size of the growth rates are different, however, depending on which methodology is used to calculate TFP. This suggests that care should be taken in interpreting the results of any productivity study, as the outcome will be affected by the measure used.

Having found evidence of some improvement in productivity since privatisation we wanted to determine how, it at all, the growth was affected by the regulatory regime. We estimated the relationship between each of our productivity growth measures and a number of regulatory and ownership variables using regression analysis. The goodness-of-fit measures for this analysis are relatively low, suggesting that other factors also have an impact on productivity growth. In addition, the estimated coefficients varied slightly in size depending on which productivity measure was used as the dependent variable. The direction of the relationships were, however, generally consistent with our expectations. The limitations of the model, and the variation across different productivity measures, makes it difficult to form any firm conclusions on the relationship between productivity growth and RPI-X regulation.

Our study of the impact of RPI-X regulation is preliminary at this stage. A number of interesting relationships have been identified, many of which conform with our ex-ante expectations. Further analysis, ideally involving an extended dataset and more explanatory variables, would be valuable in determining whether the relationships estimated here hold over a longer, and more varied, timeframe. More research is also required to determine which of the TFP measures best captures the extent of productivity improvement in the regulated sectors.

[35] Only the first methodology was used to calculate NGC's TFP growth rate.

Chapter 5

Good delegation regimes

With a price cap mechanism the regulator delegates the price choice to the firm but sets an upper limit on the price level which can be chosen[1]. The firm is therefore operating under a restricted delegation contract[2]. Delegation is used because the regulator does not know the firm's cost and is therefore unable to simply direct the firm to choose the optimal price level[3]. Holmström (1984) also emphasises that delegation has the lowest transaction costs from amongst the set of mechanisms which can be used to coordinate the firm's information.

As described in chapter 2, the regulator sets the price constraint so that average revenue is equal to the expected efficient average costs at the cap. If the firm's cost level is lower than the regulator's forecast, it is able to earn profits and there is allocative inefficiency. As discussed in Chapter 3, this arises because of asymmetric information and because the firm is provided with an incentive to undertake cost-reducing effort. The problem is particularly acute if, given its actual cost level, the firm is able to choose the monopoly price from the constrained choice set[4]. This will only arise if there is a sufficient degree of cost uncertainty - i.e. if the range of potential costs types is quite wide and the single cap is unable to capture all of the potential variation in cost levels. In addition, as discussed in further detail in section 5.1.4, a monopoly firm

[1]The chapter title is borrowed from Schmalansee (1989), 'Good Regulatory Regimes'. As in that article, we do not find the absolutely optimal regulatory mechanism but instead compare the welfare levels of four pre-defined contract options.

[2]Baron (1989) provides a useful discussion on the relationship between delegation games (where the firm makes the price decision and the regulator sets a fixed relationship between the transfer and the chosen price) and revelation games (where the firm announces its type and the regulator sets the allowed price and transfer simultaneously in a contract).

[3]Green and Stokey (1981) explain, in a more general framework, how delegation allows the principal to make better use of the agent's information. Holmström (1977) also emphasises that delegation is used 'as a means of transmitting information and overcoming informational gaps'.

[4]Laffont and Tirole (1993, p154) find that under a pure price cap regime some cost types will choose the price cap while others will choose their monopoly price.

will only price below the cap if demand is elastic. The problem of a firm being able to choose its monopoly price under a price cap regime will therefore only arise in a limited number of industries which have elastic demand and sufficient variation in expected cost levels. The analysis in this chapter may therefore not be directly relevant for the core water and electricity industries currently subject to RPI-X regulation in the UK as they face inelastic demand curves. The discussion is likely to be of more relevance for sectors in which consumers are price sensitive, perhaps because price levels are high and/or there are supply constraints due to scarcity.

Assuming a regulated industry meets the conditions of elastic demand and significant cost uncertainty, there is a concern that the regulated firm will continue to price at its monopoly level under the price cap, restricting the expected welfare benefits of this regulatory regime. This scenario raises the question of whether the regulator could increase welfare by placing more restriction on the firm's choice set, and increase the probability that the monopoly price is no longer a feasible option. The main issue is whether the extra restriction would cause the firm to choose a lower price level, thereby increasing welfare, or a higher price level. We will see that the effect depends on the regulator's belief about what the firm's cost is.

The allocative inefficiency of the price cap contract is emphasised in the theoretical literature on optimal mechanism design. Baron and Myerson (1982) examine a static model with known demand and asymmetric information about the firm's costs[5]. In this framework they found the contract which maximises expected welfare subject to a set of individual rationality and incentive compatibility constraints[6]. If the regulator's belief function about the firm's cost level satisfies the monotone hazard rate property, the optimal regulatory contract is fully separating[7]. The regulator offers the firm a menu of cost-specific price levels, and the firm chooses the price level from the menu which maximises its profit. An information rent is paid to the firm to incentivise it to

[5]Baron and Myerson's (1982) model has been extended by several other authors to include asymmetric information about demand, multi-product monopoly firms and multi-period interactions between the regulator and the firm. See, for example, Armstrong and Vickers (2000), Armstrong (1999), Baron and Besanko (1984), Finsinger and Vogelsang (1985), Laffont and Tirole (1986), Lewis and Sappington (1988a and 1988b), Loeb and Magat (1979), Riordan (1984), Sappington and Sibley (1988), Sibley (1989) and Vogelsang (2001).

[6]The revelation principle allows Baron and Myerson (1982) to focus on direct mechanisms which ensure that the firm is honest about its cost level when choosing the price level.

[7]The distribution function, $F(\theta)$, has a monotonically increasing hazard rate if:

$$\frac{\partial}{\partial\theta}\left(\frac{F(\theta)}{f(\theta)}\right) > 0$$

Laffont and Tirole (1993) suggest that this is synonomous with a 'decreasing returns assumption'. That is, the monotone hazard rate implies that the marginal gain from improved productivity decreases as the level of efficiency increases.

choose the welfare-maximising price level given its cost[8]. Each cost-type then chooses a different price level from the menu. In this analysis price cap regulation, which results in a semi-separating outcome with some types bunching at the price cap, is not optimal.

In Baron & Myerson's analysis, and throughout the mechanism design literature, termed the 'new economics of regulation' by Laffont (1994), there is an assumption that the regulator can pay an unrestricted transfer to the firm. In reality regulators are established as independent bodies which do not have the power to raise revenue to finance transfers. The regulator's only source of income is the revenue earned directly from consumers, and hence the transfer paid to the firm is limited to be less than or equal to this revenue. We assume in this analysis that the constraint holds with equality. The welfare maximising contract with unrestricted transfers is therefore not feasible in practice. This was noted by Schmalansee (1989), who emphasised that this is the primary reason why the proposed theoretical mechanisms have not emerged in practice. Laffont and Tirole (1993, Ch2) also argued that when transfers are restricted, the economic theory on optimal pricing contracts changes. We therefore expect that the welfare-maximising regulatory contract will be different when this constraint is included.

We have a situation where the contract proposed by optimal mechanism design theory is infeasible, and the price cap contract used in practice is suboptimal. The transfer restriction means that the regulator has lost some flexibility when designing the regulatory contract, and the only degrees of freedom remaining relate to the level of choice to give to the firm. This leads us to consider the question of whether an alternative delegation contract, with more or less restriction on the firm's choices than the price cap, could yield a higher level of welfare when the transfer paid to the firm is restricted.

It has proved simpler, algebraically, to determine the welfare-maximising delegation contract by considering the firm's output choice. Our discussion is therefore focused on output floors rather than price caps. We note, as confirmed by De Fraja and Iossa (1998), that fixed output regulation results in the same outcome as fixed price regulation when the firm is a single-product monopoly and demand is known by all parties. *Conclusions reached here therefore immediately apply to the situation where a regulator sets restrictions on the firm's price decision.* In addition, an analysis based on output decisions allows for direct comparisons with the existing literature on optimal regulatory mechanisms which tends to focus on output rather than price choices.

In this chapter we abstract from the detail of how RPI-X regulation works in prac-

[8]A useful exposition of the rent extraction-efficiency trade-off in principal-agent models can be found in Laffont and Martimort (2002).

tice, and use a simplified framework to analyse the question of whether the output floor contract is the best means of regulating the monopoly firm. The welfare properties of a range of output contracts are examined using a single-period model where the monopoly firm's cost level lies in a bounded continuum. The regulator does not know the firm's cost level but he does know the bounds of the continuum of costs, and he has a belief function over the continuum about where the firm's actual costs lie. The interval of possible cost types is wide, reflecting the regulator's uncertainty about the firm's efficiency level. The regulator and firm have full information about the elastic demand function. In addition, the regulator has no outside source of financing, and the transfer paid to the firm is restricted to be equal to the revenue earned from consumers.

With this framework we compare the welfare level under each of the following contracts.

1. An output floor contract in which the regulator delegates the output choice to the firm but restricts the firm's choice to be above a fixed point (the floor). The firm can choose any output level on the continuous interval above the floor.

2. A welfare-maximising fully separating contract, along the lines of that proposed by Baron and Myerson (1982), in which the firm chooses its output level from a menu of cost-specific outputs designed by the regulator. It turns out that the outcome with this contract corresponds to the case of unrestricted delegation.

3. A no-delegation contract in which the firm is offered a take-it-or-leave-it single output level. The output level is equal to the welfare-maximising pooling output level.

4. A restricted two-interval delegation contract in which the firm is offered a choice set which is equal to that offered under the output floor contract, but with an extra set of output levels now disallowed. That is, the regulator introduces a discontinuity into the set of allowed output levels relative to the output floor contract, and thereby reduces the amount of choice which the firm has.

We emphasise that we are not designing the absolutely optimal contract but are instead finding the *best* contract from amongst this set of feasible contracts. That is, we are assessing whether one form of contract dominates another from a given set of options. In section 5.1 we describe the model in more detail. In section 5.2 each contract type is derived, and the firm's choices under that contract are explained. Welfare under the unrestricted delegation and no delegation contracts are compared to welfare under the output floor contract in section ??. We find that, when the transfer is restricted, the output floor contract yields a higher level of welfare than both these extremes. That is, delegation with a lower bound on the firm's choice set is better than no delegation and unrestricted delegation.

In section 5.4 we explore the question of how much restriction should be placed on the firm's choice set by comparing the output floor contract to the restricted two-interval delegation contract. We find that the output floor contract, despite its prevalence in practice, is not always best. Given this, we derive sufficient conditions under which the regulator is better off placing an additional restriction on the firm's choice set, and sufficient conditions under which the regulator is better off with the output floor contract. The analysis is extended, in section 5.5, to include contracts with even more restrictions than the two-interval contract. We use a simulated example to examine how an increase in the level of restriction affects the level of welfare. Section 5.6 concludes with a summary of our main findings and a discussion of possible extensions to the analysis.

Before undertaking our own analysis we briefly review the relevant literature in this area. The literature on price cap regulation does not make explicit comparisons between the rules observed in practice and those which have been derived from first principles in the mechanism design literature. Instead this body of literature focuses on comparing the efficiency properties of existing regulatory rules to other existing rules, as epitomised by the volumes of research comparing price cap regulation to rate of return regulation[9]. There has therefore been no formal analysis comparing the contracts proposed by theory and those which are used in practice. Our contributions to the literature are the *explicit exposition* of how the transfer restriction affects the optimal direct mechanism originally derived by Baron and Myerson (1982), and the formal comparison of the welfare properties of the price cap contract with the optimal fully separating and pooling direct mechanisms. These comparisons are of interest for industries with elastic demand and significant cost uncertainty.

Laffont (1994) presents a brief analysis of the impact of restricted transfers on the mechanism design problem when the firm can alter its costs through investment in cost-reducing effort. He finds that when transfers are restricted, and costs are unobservable, high rents are required to provide optimal efficiency incentives to the firm. A fully separating contract will therefore be costly for the regulator. Levaggi (1999) considers the optimal level of procurement by the government when it faces a stringent budget constraint. The conclusion of this analysis is that a fully separating contract will not

[9]See, for example, Bradley and Price (1988 and 1991), Braeutigam and Panzar (1993), Crew (1994), Joskow and Schmalensee (1986), Littlechild (1983), Lyon (1996), and the RAND (1989) *Symposium on Price Cap Regulation* . Laffont and Tirole (1993, Sections 2.6 and 2.7) provide a useful guide to the types of mechanisms that are considered. Gasmi, Ivaldi and Laffont (1994) provide an evaluation of many of these regulatory schemes using simulation techniques.

be allocatively efficiency because it is 'too expensive for the principal'. The optimal contract involves a pooling equilibrium or a bargaining solution. Restricted delegation contracts, including price cap regimes, which result in semi-separating outcomes are not explicitly considered in these papers.

The question of how much restriction to place on the firm's choice set has also not been considered in detail in the delegation literature. Papers on optimal delegation in *static models* tend to presume that the choice set is made up of a continuous interval, with a lower and upper bound at most[10]. For example, Holmström (1977,1984) found that there is an optimal solution to the principal's delegation problem which comprises a nondegenerate and finite single interval. In his analysis of quantity controls, where upper limits are placed on the amount a manager can produce, he found that the principal will prefer an interval without a gap to one with a gap when the state variable has a normal distribution. This result is specific to the conditions of the example however. Indeed Holmström (1977) recognised that there may be situations where the single interval is not optimal but argues that 'If the optimal control has gaps, it is quite hard to derive, and certainly an analysis of interval controls can be defended on pragmatic grounds'.

Despite this, there are examples of dynamic delegation models, with moral hazard and asymmetric information, which find that the optimal choice set is disjoint (ie, made up of more than one interval). For example Szalay (2001) finds that the principal may wish to curtail the agent's decision set so that there is a region of 'prohibited intermediate choices' as well as a region of 'prohibited extreme choices'. In line with this we argue that, despite the increased complexity involved, regulators should explore the option of restricting the firm's choices further if there is evidence that this will increase welfare. The need for alternative contracts is only relevant, however, in industries with elastic demand and where there is a wide range of potential cost types (i.e. cost uncertainty). The analysis of the preferred level of restriction to place on the firm's choice introduces a new aspect to the design of regulatory mechanisms, and opens up a new set of issues for regulators to consider in practice.

[10]Armstrong (1994) does not simply assume that the continuous interval contract is optimal but finds, in his Appendix A, local sufficient conditions for this to be the case. He concludes that if the principal's beliefs about the state of the world are such that the probability density function is relatively flat (a more precise condition is provided in the article) the single interval decision set is locally optimal. As our objective function differs from that analysed by Armstrong (for example it is not always concave in the agent's action choice), and we do not impose any prior restrictions on the shape of the regulator's belief function, we expect to find a different set of conditions. We note that in later versions of this working paper Armstrong simply assumes that the decision set will be a continuous interval and does not analyse when this is optimal.

5.1 The Model

We use a stylised single period model to compare the level of welfare under different regulatory contracts. The key features of the model are outlined below. All functions are assumed to be continuous and differentiable.

5.1.1 The firm

The firm is the sole provider of a single product to consumers. There is a unit price for the product. Price and output are observable and verifiable. There is no threat of entry as the firm possesses an indefinite licence to operate as a monopolist in the market.

The firm faces a linear *demand function*:

$$q(p) = a - bp \quad a > 0, b > 0$$

We assume that there is always some production, $q(p) > 0$. The demand function is known with certainty by both the regulator and the firm. As part of its licence obligation the firm is required to always meet demand.

The firm's production costs are:

$$C(\theta) = \theta q$$

The *marginal cost* of producing network access, θ, represents the firm's efficiency type and it can take on any value in the continuous interval $[\theta_L, \theta_H]$. The firm is classified as efficient if it has a low marginal cost, θ_L, and as inefficient if it has a high marginal cost, θ_H. The firm's cost level is exogenous. In particular the firm cannot change its cost level by investing in cost-reducing effort. We assume that the firm knows its own type with certainty but the regulator does not know the firm's type. The model is therefore one of adverse selection only, rather than moral hazard and adverse selection[11].

We assume that the firm earns all of its *revenue, T*, from the regulator. We use the standard accounting convention that consumers pay the regulator for the product consumed, and the regulator then chooses how much income to transfer to the firm. That is, the regulator acts like a clearing-house between consumers and the firm[12].

[11]Laffont and Martimort (2002) emphasise that in models with cost-reducing effort the ex-post observability of cost technically makes the single-period model with effort a model of adverse selection only anyway.

[12]A similar approach is adopted by Laffont and Tirole (1993) where they assume that the revenue 'generated by the sale of the outputs is received by the state and that the cost is reimbursed to the firm' (p168).

This simplifies the analysis greatly while having no impact on the firm's incentives relative to the case where it earns revenue directly from consumers.

The firm chooses the output level which maximises profit. The profit function is:

$$\pi(\theta) = T - \theta q$$

5.1.2 The regulator

The regulator makes a take-it-or-leave-it contract offer to the firm. The offer specifies the type of regulatory contract which the regulator intends to use and the precise details of that contract. We discuss, in section 5.2, the contract options which are considered by the regulator. We assume that both parties can commit to the terms of the contract and that the legal framework for enforcing the contract exists. This means that once the contract has been agreed, and the firm's output choice is revealed, the regulator cannot change the regulatory mechanism.

The regulator knows the demand function with certainty but does not know the firm's cost type. We assume that he has a prior belief that the firm's efficiency level is distributed with a distribution function $F(\theta)$ on the interval $[\theta_L, \theta_H]$. The associated probability density function is $f(\theta)$. We assume that the regulator's belief coincides with the actual distribution of types (ie, his belief structure is accurate) and that the probability distribution is common knowledge. We also assume that the density function is strictly positive for all types and, hence, the distribution function is continuous and increasing.

Objective function

Following Baron and Myerson (1982) we assume that the regulator maximises a weighted sum of consumer welfare (CW) and the firm's profits ($\pi(\theta)$):

$$W(q, \theta) = CW + \lambda\pi(\theta)$$

The weight on profits, λ, lies in the interval $(0, 1)$. As noted by Laffont and Tirole (1993), this weight is synonymous with the regulator discounting the firm's welfare relative to consumer welfare. As the weight is less than one the regulator cares more about consumers than the firm and therefore dislikes providing the firm with a positive rent.

Gross consumer surplus for the known demand function is:

$$V(q) = \int_o^q p(x)dx = \frac{aq}{b} - \frac{q^2}{2b}$$

Consumer welfare is equal to gross consumer surplus minus the revenue paid to the firm:

$$CW = \frac{aq}{b} - \frac{q^2}{2b} - T$$

Given uncertainty about the firm's type the regulator's objective function is:

$$\begin{aligned} E_\theta(W(q,\theta)) &= \int_{\theta_L}^{\theta_H} W(q,\theta) f(\theta) d\theta \qquad \textbf{Expected Welfare} \\ &= \int_{\theta_L}^{\theta_H} \left\{ \frac{aq}{b} - \frac{q^2}{2b} - \lambda\theta q - (1-\lambda)T \right\} f(\theta) d\theta \end{aligned}$$

Constraints

When maximising expected welfare the regulator faces two standard mechanism design constraints - an individual rationality constraint and an incentive compatibility constraint. We also introduce a third constraint on the transfer paid to the firm.

Under the *individual rationality* constraint the regulator ensures that the firm makes non-negative profits[13]:

$$\pi(\theta) \geq 0 \ \forall\theta \qquad \textbf{Rationality}$$

This implies that the regulator must ensure that the firm, whatever its type, continues to produce under the regulatory contract (ie, there is no firm shut-down).

Under a direct mechanism the firm, of type θ, announces its type to be $\widehat{\theta}$. The regulator presents the firm with a contract $\left\{T(\widehat{\theta}), q(\widehat{\theta})\right\}$ and the firm earns profit:

$$\pi(\theta, \widehat{\theta}) = T(\widehat{\theta}) - \theta q(\widehat{\theta})$$

The *incentive compatibility* constraint ensures that the firm's announcement is honest. It is stated as follows.

$$\pi(\theta) = \max_{\widehat{\theta}} \pi(\theta, \widehat{\theta}) \ \forall\theta, \widehat{\theta} \qquad \textbf{Compatibility}$$

Under this constraint it is privately optimal for the firm to report its type honestly[14].

Finally we assume that the regulator has no outside source of income - eg, unrestricted money from the Treasury or the right to impose taxes directly on consumers - and hence the amount of revenue which is transferred to the firm is restricted to be less

[13] We have an 'interim individual rationality constraint' because we are considering the case where the contract is agreed after the firm knows its type.

[14] This constraint is only relevant when the regulator offers a menu of type-specific contracts. In this situation a firm may have a private incentive to misrepresent its type and the regulator designs the contract to offset this incentive. When non-type specific pooling and delegation contracts are offered the firm gets the same contract offer no matter what its announced type and hence incentive compatibility is not an issue.

than or equal to the revenue earned from consumers. We assume that the constraint holds with strict equality[15]:

$$T = pq(p) = \frac{aq - q^2}{b} \qquad \textbf{Restricted transfer}$$

This restriction reflects the idea that the regulator is an independent body which has no access to Treasury funds and which is not required to provide funds to the Treasury. Utility regulators in the UK were created with this kind of independent status. We recognise, at the outset, that the regulator would be better off if the financing was available for an unrestricted transfer but we assume that the existing institutions and legislation are unlikely to change in practice and, hence, transfers are likely to remain restricted for the foreseeable future[16]. In comparison to the mechanism design literature with unrestricted transfers, the regulator has lost an important regulatory instrument and we therefore expect him to have less degrees of freedom when choosing the welfare-maximising contract. Indeed the only instruments available to the regulator are the form of the contract and the level of output(s) allowed in that contract[17].

With the transfer restriction imposed the regulator's objective function becomes:

$$E_\theta(W(q,\theta)) = \int_{\theta_L}^{\theta_H} \left\{ \frac{(1-2\lambda)q^2}{2b} + \frac{\lambda (a - b\theta)\, q}{b} \right\} f(\theta) d\theta$$

Appendix F considers the comparative static properties of this function. The main results are as follows.

- The level of welfare is increasing in output for a firm of type θ (Lemma F.1).
- The rate of change in welfare, arising from a change in output, is increasing in the level of output if the welfare function is convex ($\lambda \in (0, 0.5)$) (Lemma F.2).
- The rate of change in welfare, arising from a change in output, is decreasing in the level of output if the welfare function is concave ($\lambda \in (0.5, 1)$) (Lemma F.3).
- Welfare is decreasing in type for a given level of output. The rate of decrease is constant for all types (Lemma F.4).
- The rate of change in welfare is decreasing in type for a given level of output. The rate of decrease is constant for all types (Lemma F.5).

[15]We also rule out the possiblity that the regulator sets a two-part tariff. This pricing option has been used by some authors, for example Baron and Besanko (1987), to mirror the role of an unrestricted transfer function. The fixed payment in the two-part tariff, which is funded by consumers, is used to influence the firm's incentives in the same way as a transfer.

[16]Laffont and Tirole (1993, Ch15) explain how concerns about the regulator misusing his power, perhaps because of regulatory capture, provide a rationale for not allowing the regulator to control an unrestricted transfer budget.

[17]The form of the contract relates to the amount of restriction which is placed on the firm's choice set.

5.1.3 The output floor contract (Contract F)

We assume, initially, that the regulator limits the monopoly firm's behaviour by using an output floor contract. In subsequent sections we compare welfare under this contract to the welfare level under a number of alternative contracts. Under the output floor contract the regulator delegates the output choice to the firm but restricts the firm's choice to lie in a single interval which is bounded below by the output floor[18]. The contract is:

$$\text{choose } q \epsilon \left[q_{\min}, \infty\right]$$

The set of forbidden output levels is: $[0, q_{\min})$.

The regulator must ensure that the firm continues to produce under this regulatory contract (individual rationality). A sufficient condition for this is that: $\pi(\theta_H) \geq 0$. The minimum feasible level of output is therefore set where the inefficient type's profits are exactly equal to zero. This gives us the output floor:

$$q_{\min} = a - b\theta_H$$

For notational simplicity we continue to use the term $q_{\min}$, recognising this restriction on its value.

The conditions of the contract restrict the firm's choice from below. We know that as a profit maximiser the firm will produce its monopoly output level if it is within its allowed choice set. If, however, the monopoly output level is not available, the firm will choose the level of allowed output which is closest to it. This is ensured by the fact that the firm's profit function, incorporating the transfer restriction, is concave and symmetric in output. This gives us:

$$\begin{aligned} q &= q^M(\theta) \text{ if } q^M(\theta) \ \epsilon \left[q_{\min}, \infty\right] \\ q &= q_{\min} \text{ if } q^M(\theta) \ < q_{\min} \end{aligned}$$

The semi-separating choices under the output floor contract are illustrated in figure 5.1 and summarised as:

$$\begin{aligned} q &= q^M(\theta) \text{ if } \theta \epsilon \left[\theta_L, \frac{a - 2q_{\min}}{b}\right] \\ q &= q_{\min} \text{ if } \theta \epsilon \left(\frac{a - 2q_{\min}}{b}, \theta_H\right] \end{aligned}$$

[18] The regulator is only concerned with types who produce output levels that are too low. It is therefore presumed unnecessary to place on upper limit (ie, ceiling) on the firm's output level. We note that the firm optimally chooses never to produce more than $q(\theta) = a - b\theta$, as any higher output level results in a loss.

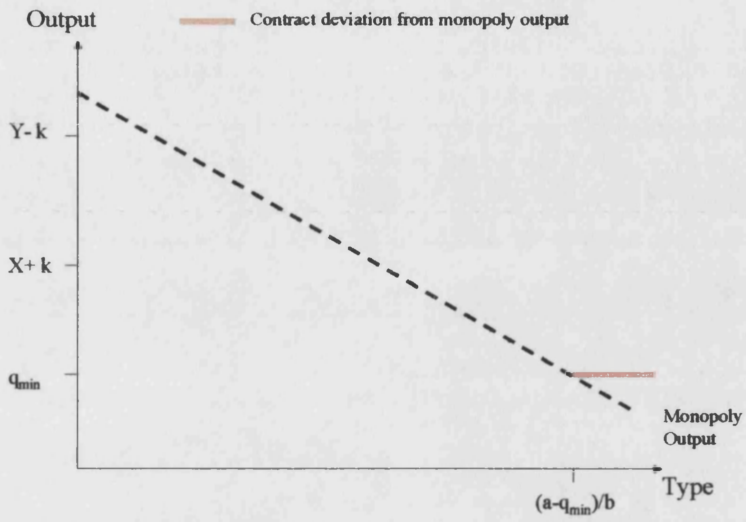

Figure 5.1: Output choices under contract F

5.1.4 Industry characteristics

In our description of the output floor contract we found that some cost types would be able to choose their monopoly output level as it lies above the output floor. That is, for some types:

$$a - b\theta_H < \frac{a - b\theta}{2} = q^m(\theta)$$

We consider here the conditions which need to prevail in the regulated industry for this outcome to arise. *When comparing the output floor contract to other contracts we assume that these characteristics exist.*

Cost constraints

The firm's monopoly output level will be greater than the output floor if:

$$\begin{aligned} \frac{a}{b} &< 2\theta_H - \theta \ \forall\theta \\ \Rightarrow \frac{a}{2b} + \frac{\theta}{2} &< \theta_H \ \forall\theta \end{aligned}$$

As this condition must also hold for θ_L we find that the interval of types must satisfy the following constraint if some types are able to choose their monopoly output level above the output floor:

$$\frac{a}{b} - \theta_H < \theta_H - \theta_L$$

The range of possible cost levels must therefore be sufficiently wide - i.e. there must be sufficient uncertainty about the cost level. *This is a necessary and sufficient condition for the situation to arise where some types will be able to choose their monopoly output level under the output floor contract.*

In addition, if the firm prices at marginal cost, the level of output should be positive. This will always arise if costs satisfy the condition:

$$0 \leq \theta_L < \theta_H < \frac{a}{b}$$

This ensures that:

$$0 < \frac{a}{b} - \theta_H < \theta_H - \theta_L$$

Elasticity of demand

The elasticity of demand in the regulated industry is:

$$\varepsilon_p = \frac{-bp}{a - bp}$$

At the output floor the elasticity is equal to:

$$\varepsilon_p = \frac{-b\theta_H}{a - b\theta_H}$$

The demand curve is elastic at the output floor if:

$$\begin{aligned} \left|\frac{b\theta_H}{a - b\theta_H}\right| &\geq 1 \\ \Rightarrow \quad \theta_H &\geq \frac{a}{2b} \end{aligned}$$

If the demand curve is elastic, a firm of type θ will be able to choose its monopoly output level under the output floor contract if::

$$\frac{a}{2b} \leq \frac{a}{2b} + \frac{\theta}{2} \leq \theta_H$$

However, the monopoly output level will not be feasible under the output floor contract for any type, even when demand is elastic, if:

$$\frac{a}{2b} \leq \theta_H \leq \frac{a}{2b} + \frac{\theta_L}{2} < \frac{a}{b}$$

Elasticity of demand is therefore consistent with the monopoly output level being above the output floor in some circumstances but it is not a sufficient condition. The constraints on the cost interval must also be met.

The demand curve is inelastic at the output floor if:

$$\theta_H < \frac{a}{2b}$$

With the inelastic demand curve the constraint on the width of the cost interval is never satisfied. This is because the inelasticity constraint at the high cost level, and the constraint on the width of the cost interval, can not be met simultaneously since:

$$\frac{a}{2b} + \frac{\theta}{2} < \theta_H \ < \frac{a}{2b} \forall \theta \text{ is never feasible given } \theta > 0$$

With an inelastic demand function no type will be able to choose its monopoly output level under the output floor contract. This suggests that *elasticity of demand is a necessary but not sufficient condition for the situation to arise where a regulated firm is able to choose its monopoly output level under the output floor contract.* Therefore, before deciding whether or not to change the form of the regulatory contract, the regulator will first need to calculate the elasticity of demand for the particular industry in question. This will allow him to determine whether the concern discussed in this paper arises in that particular industry.

Industries of interest

We conclude that a firm *may* be able to choose its monopoly output level under the output floor contract if:

- demand is elastic; and
- there is a sufficiently wide interval of cost types (i.e. a high degree of cost uncertainty).

Table 5.1 provides estimates of the own price elasticity of demand in a range of utility industries across a number of countries[19]. These show elasticity estimates less than one for electricity and water, supporting the generally accepted conclusion that demand in these sectors is relatively inelastic. The inelasticity is largely affected by the necessary nature of these products, with consumers being unwilling to reduce consumption in response to a price increase. Furthermore, particularly in the case of water, the absence of alternative substitutes affects the degree to which consumers can respond to price changes. The estimates for the mobile telecommunications industry are highly variable, with evening demand close to being elastic. However there is

[19]We note that the methodologies and data used to measure these demand elasticities are often subject to criticism. It is beyond the scope of this paper to critically assess the estimates but we accept that they may not be robust. They are used here purely for illustrative purposes.

evidence that up to 22% to 29% of consumers in the UK switch operators in search of a cheaper tariff package, suggesting that consumers are highly price sensitive in this sector[20]. Demand for natural gas appears to be elastic.

These aggregate measures may conceal the fact that some consumers - notably large users - will be more price sensitive than others. For example, if the price of water increases, a factory which uses large volumes of water on a daily basis - and is metered - will use water more efficiently by introducing new technologies and working practices. Similarly a large user of gas will become more energy efficient in response to a price increase so as to reduce the total cost to the firm. In addition, electricity supply businesses may encourage their end-user customers to be more energy efficient to reduce the overall costs of electricity purchased, thereby making the demand which electricity distribution companies face more elastic. Regulated businesses which have a higher proportion of large user customers may therefore have a more elastic demand function.

Consumers are more likely to be price sensitive if water and energy resources become scarce - placing pressure on consumers to reduce the volumes used - and if the cost of water or energy is a high proportion of a company or household's overall expenditure[21]. In some countries water efficiency schemes are introduced which actively encourage consumers to be more price sensitive and to reduce the level of water used. For example, in California and New York water prices were high because of scarcity problems and households purchased new water efficient technologies and changed their consumption behaviour in response to these prices. This again emphasises the point that, when considering the question of whether or not to change the form of the regulated contract, a regulator should estimate the elasticity of demand for the particular regulated firm as the actual elasticity for a sector may vary by firm, by country and/or over time.

We conclude that the concern that a regulated firm will be able to choose its monopoly output level under the output floor contract is less likely to arise in the regulated water and electricity distribution sectors in the UK than in other sectors such as gas and mobile telecommunications. This is because current estimates indicate that demand, at least at an aggregate level, is currently inelastic in these sectors. Further estimates of the actual elasticity of demand for individual firms, taking account of the

[20]Competition Commission (2002, p53, para 2.193), *Vodafone, O2, Orange and T-Mobile: Reports on references under section 13 of the Telecommunications Act 1984 on the charges made by Vodafone, O2, Orange and T-Mobile for terminating calls from fixed and mobile networks.*

[21]For example, water companies are large users of electricity and gas. As energy prices increase these companies look for ways to become more energy efficient. Similarly, the electricity generating companies are large users of water and also seek to implement water efficient technologies as water prices increase.

characteristics of their consumers, would be required to confirm that this is the case. This analysis is beyond the scope of this paper but would be a useful next step in future research.

The analysis in this chapter may therefore not be highly relevant for the core regulated sectors in the UK. The problem discussed here may arise in other sectors where consumers are more price sensitive. Furthermore the problem may arise in the regulated network utility sectors in other countries - which face different supply constraints and price levels - and may even become more relevant in the UK if the nature of demand changes over time or if firm-specific elasticity estimates reveal more price sensitivity than the aggregate historical estimates suggest.

Table 5.1: Estimates of own-price elasticity across a range of sectors

Industry/Product	Own price elasticity
UK fixed to mobile calls (1997 to 2001)[1]	
- daytime	-0.42 to -0.33 (long run)
	-0.26 to -0.17 (short run)
- evening	-0.95 to -0.76 (long run)
	-0.58 to -0.41 (short run)
- weekend	-0.54 to -0.43 (long run)
	-0.20 to -0.20 (short run)
Energy demand	
Australian residential energy demand (1984 ro 1998)[2]	
- electricity	−0.35 to -0.34
- natural gas	−1.03 to -1.02
Norwegian household electricity demand (1976 to 1993)[3]	
Short run	-1 to -0.4
Long run	-1 to -0.3
US urban residential energy demand (1972 to 1975)[4]	
Electricity	-0.25 to -0.72
Gas	−2.07 to -0.94
UK electricity demand (1937 to 1938)[5]	-0.89 (short run)
Price elasticities for public water supply[6]	
Australia Sydney Water (1960 to 1994)	-0.13
Copenhagen, Denmark (early 1990s)	-0.10
France (1988 to 1993)	-0.26 to -0.17
Italy (mid-1990s)	-0.24
Korea (1998)	-0.29
New Zealand (1980s)	-0.29 to -0.08
Illinois, US (mid-1980s)	-0.71 to -0.48

Sources: [1]dotecon (2001); [2] Akmal and Stern (2001); [3]Halvorsen and Larsen (1999); [4]Lakshmanan and Anderson (1980); [5]Houthakker (1951); [6]OECD (1999)

5.2 Alternative Contract Options

We compare expected welfare under the output floor contract to welfare under three other contract options. **We assume that the regulated industry has an elastic demand curve and that the expected interval of costs satisfies the conditions described in section 5.1.4.** There is thus a risk that the regulated firm will be able to choose its monopoly output level under the output floor contract. The objective is to determine whether the contract with a single constraint, which is observed frequently in practice, is in fact 'best'[22]. If welfare is found to be higher under an alternative contract option, this raises the question of whether existing regulatory mechanisms - such as the RPI-X regime - should be changed.

Given asymmetric information about the firm's costs, the regulator is unable to set a direct output rule for the firm and first-best is not attainable[23]. We therefore focus our attention on a number of second-best feasible mechanisms which take account of the asymmetric information in the model. Specifically, we assume that the regulator can offer the firm one of the following alternative regulatory contracts. We explain below how the regulator determines each contract and how the firm makes choices under the contract.

- *Optimal fully separating contract* ***(Contract S)*** - the regulator offers the firm a menu of type-specific output levels and the firm chooses one output from this menu of discrete points. The contract is:

$$\text{choose } q \epsilon \{q(\theta)\}$$

  The regulator's menu of type-specific output levels is determined by imposing full separation and, given this, maximising expected welfare subject to the individual rationality, incentive compatibility and transfer constraints. As we will show this contract can be implemented by simply delegating the output decision to the firm without placing any restrictions on the choice set. It therefore corresponds to the case of *unrestricted delegation* and there are no forbidden output levels.

- *Optimal pooling contract* ***(Contract P)*** - the regulator offers the firm a single take-it-or-leave it output-level, q^P. The output level is determined by imposing the condition that the contract is pooling and, given this, maximising expected

[22]We reiterate that, in the spirit of Schmalensee (1989), we are seeking the contract which yields the highest level of welfare from amongst a predetermined set rather than the absolutely optimal contract.

[23]This is the level of output which maximises the unrestricted social welfare function, when the firm's type is known, subject to the constraint that the firm make non-negative profits. For a firm of type θ the first-best output level is $q^{FB}(\theta) = a - b\theta$.

welfare subject to the individual rationality and transfer constraints. This contract corresponds to the case of *no delegation*. All output levels except one are forbidden.

- *Restricted two-interval contract* (**Contract** R^2) - the regulator delegates the output choice to the firm but restricts the firm's choice set. The restricted choice set is the same as that in the output floor contract but with an additional set of output levels disallowed. The contract is:

$$\text{choose } q\epsilon\,[q_{\min}, X+k] \cup [Y-k, \infty]$$

The set of *extra* forbidden output levels, relative to the output floor contract, is: $(X+k, Y-k)$. X and Y are determined exogenously and they satisfy the condition that $0 \leq q_{\min} < X < Y$[24]. The parameter k determines the size of the interval of additional disallowed output levels, with $0 \leq k < \frac{Y-X}{2}$. *The smaller is k, the larger is the set of disallowed outputs.* The output floor contract is an extreme case of this contract where $k = \frac{Y-X}{2}$.

The options presented here reflect different degrees of choice given to the firm. They range from unrestricted delegation of the output choice to complete restriction (or no delegation). In between these extremes the regulator can place intermediate levels of restriction on the firm's choice set. We focus initially on the restricted two-interval contract, but consider the case for contracts with more than two-intervals in section 5.5. The greater the number of intervals, the more restriction is placed on the firm's choice. This is because the set of disallowed output levels has increased relative to the output floor contract. This analysis allows us to determine what the *optimal form* of the delegation contract is[25]. Once the welfare-maximising contract form has been determined the regulator will calculate the optimal value of the interval boundaries in this contract. This issue is not considered in this paper but is clearly the next required step in future research.

5.2.1 Optimal fully separating contract (Contract S)

We determine the optimal fully separating and pooling contracts by extending Baron and Myerson's (1982) mechanism design model to include our restriction that the

[24]The determination of these boundaries is not considered in this paper although it is clear that for the bounds to have an impact on the firm's choices we require $X < Y < q^m(\theta_L)$. $q^m(\theta_L)$ is the monopoly output level of the efficient firm and is equal to $\frac{a-b\theta_L}{2}$.

[25]The reader is referred to Holmström (1984) for the theorem on the existence of a solution to the delegation problem.

transfer paid to the firm is equal to the revenue earned from consumers. This additional constraint changes Baron and Myerson's results significantly.

The regulator's problem is to determine the contract which will maximise expected welfare subject to the transfer restriction, the individual rationality constraint and the incentive compatibility constraint. It can be stated as follows:

$$\max_{\{T(\theta),q(\theta)\}} \int_{\theta_L}^{\theta_H} \left\{ \frac{aq(\theta)}{b} - \frac{[q(\theta)]^2}{2b} - \lambda\theta q(\theta) - (1-\lambda)T(\theta) \right\} f(\theta)d\theta$$

subject to:

$$\pi(\theta) \geq 0 \ \forall\theta \qquad \textbf{Rationality}$$

$$\pi(\theta) = \max_{\widehat{\theta}} \pi(\theta, \widehat{\theta}) \ \forall\theta, \widehat{\theta} \qquad \textbf{Compatibility}$$

$$T(\theta) = \frac{aq(\theta) - [q(\theta)]^2}{b} \qquad \textbf{Transfer}$$

Assuming that the profit function is differentiable incentive compatibility implies, at the optimum where $\widehat{\theta} = \theta$, that:

$$\frac{\partial\pi\left(\theta, \widehat{\theta}\right)}{\partial\theta} \Big|_{\widehat{\theta}=\theta} = 0$$
$$\Rightarrow \ T'(\theta) = \theta q'(\theta)$$

The restriction on the level of transfers implies that:

$$T'(\theta) = \frac{aq'(\theta) - 2\left(q(\theta)\right)q'(\theta)}{b}$$

Therefore, if both the incentive compatibility constraint and the restricted transfer constraint are to be satisfied we require:

$$\theta q'(\theta) = \frac{[a - 2\left(q(\theta)\right)]\, q'(\theta)}{b} \qquad \textbf{Feasibility}$$

In Baron and Myerson's (1982) paper, and in much of the literature which has stemmed from it, there is an assumption that the regulator's belief function satisfies a monotone hazard rate property[26]. This ensures, in the analysis when transfers are unrestricted, that the optimal contract is always fully separating. Here, however, the

[26] With a monotone hazard rate the conditional probability of there being no further efficiency improvements increases as the firm's type increases (ie, as the firm becomes more efficient, the probability of it making further efficiency improvements decreases). That is, as the firm gets closer to the efficiency frontier a marginal improvement in efficiency becomes more difficult. Bagnoli and Bergstrom (1989) provide details of the many standard distributions which satisfy this condition. They also provide sufficient conditions on the density function for the monotone hazard rate property to hold.

condition for incentive compatibility and the transfer restriction to hold simultaneously is *independent of the regulator's belief function* and, hence, the monotone hazard ratio assumption no longer ensures that the optimal contract is fully separating. Indeed, a contract which satisfies this feasibility condition might be fully separating, pooling or semi-separating.

Full separation implies that output cannot be constant over an interval of types ($q'(\theta) < 0$). The feasibility condition is then only satisfied if:

$$q(\theta) = \frac{a - b\theta}{2}$$

Only this output level ensures that both constraints are satisfied. Notably this is the *monopoly level of output* and it is the only feasible contract which is available to the regulator if the regulatory mechanism is forced to be fully separating. This is consistent with Vickers and Yarrow's (1988a) prediction that, when lump-sum transfers are not feasible, 'Price would have to exceed unit cost' and there would be a further 'departure from allocative efficiency'. The outcome is also an extreme confirmation of Laffont and Tirole's (1993, Ch2) claim that the transfer restriction will lead to 'an increase in the consumer price away from marginal cost and toward the monopoly level'. We therefore conclude that with this contract there is no evident benefit of regulation, and the outcome is consistent with the unrestricted delegation of the output choice to the firm.

The output choices under the optimal fully separating contract are illustrated in Figure 5.2 and summarised as:

$$q = q^M(\theta) \ \forall \, \theta$$

We note that Baron and Myerson's (1982) level of output under the optimal contract with unrestricted transfers is:

$$q^*(\theta) = a - b\left[\theta + (1 - \lambda)\frac{F(\theta)}{f(\theta)}\right]$$

The introduction of the transfer restriction alters the optimal contract in the following ways.

- Our optimal regulatory contract might be fully separating, pooling or semi-separating, even if the regulator's belief function satisfies the monotone hazard rate property. Baron and Myerson's optimal contract is always fully separating with this assumption on the regulator's belief function.

- With our fully separating mechanism all types produce the monopoly level of output (ie, less than the first-best level). Some types are producing more than they do under Baron and Myerson's contract while others are producing less.

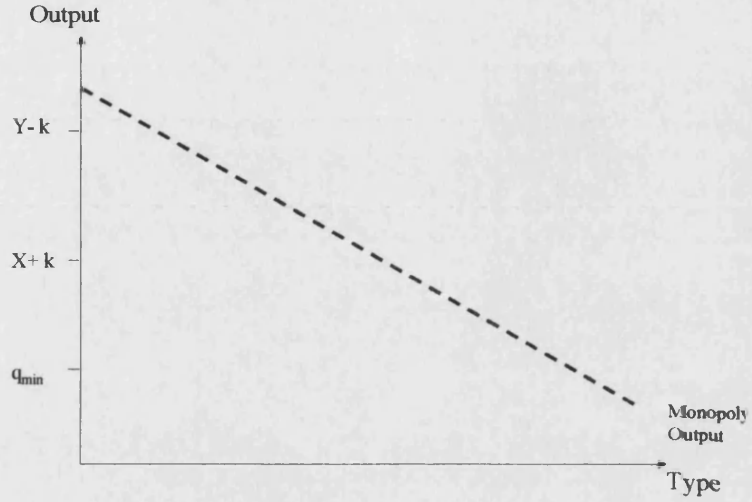

Figure 5.2: Output choices under Contract S

The net welfare effect of the transfer restriction depends on the shape of the regulator's belief function over cost types.

- The role of the transfer as a means of providing the firm with an incentive to be honest is eliminated. Instead the regulator has to allow all types to earn high profits - the monopoly level - to ensure incentive compatibility.

5.2.2 Optimal pooling contract (Contract P)

The optimal pooling contract is a special case from the set of contracts which satisfy the above constrained welfare maximisation problem. The feasibility condition is trivially satisfied as the output level is constant across all types ($q'(\theta) = 0$). The regulator only needs to consider the impact of the individual rationality and transfer constraints on the welfare maximisation problem. The regulator's problem becomes:

$$\max_{\{T^p, q^p\}} \int_{\theta_L}^{\theta_H} \left\{ \frac{aq^p}{b} - \frac{[q^p]^2}{2b} - \lambda\theta q^p - (1-\lambda)T^p \right\} f(\theta) d\theta$$

subject to:

$$\pi(\theta) \geq 0 \ \forall \theta$$

$$T^p = \frac{aq^p - q^{p2}}{b}$$

To minimise the cost of ensuring individual rationality for all types, the regulator sets the transfer so that the profit of the highest cost firm is equal to zero. We then have: $T^p = \theta_H q^p$. This condition is only consistent with the transfer restriction when:

$$\begin{aligned} q^p &= a - b\theta_H \\ T^p &= a\theta_H - b\theta_H^2 \end{aligned}$$

Notably the output level is again independent of the regulator's belief function. The most inefficient type produces the first-best level of output and makes zero profits. All other types produce an output level that is lower than first-best, and make positive profits equal to $\pi(\theta) = (a - b\theta_H)(\theta_H - \theta) > 0$. We note that in this contract all firms produce at the equivalent level of output as the output floor in contract F. This is not surprising given that it is the individual rationality constraint which influences the contract design in both cases.

The choices under the optimal pooling contract are shown in Figure 5.3 and summarised as:

$$q = q_{\min}\ \forall\ \theta$$

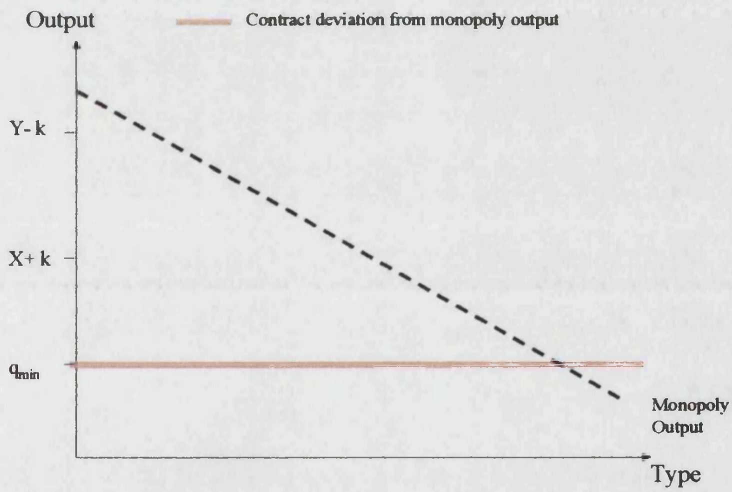

Figure 5.3: Output choices under Contract P

5.2.3 Restricted two-interval contract (Contract R^2)

In our final contract the regulator takes the optimal output floor contract and introduces an additional set of disallowed output levels in the firm's choice set. This is thus an alternative delegation contract but with increased restriction relative to the output floor contract. The profit-maximising firm will choose its monopoly output level if it lies within the allowed choice set. If the monopoly output level lies outside the firm's choice set, it will choose the closest allowed output level given the concavity and symmetry of the profit function.

The optimal output choices, as a function of type, are illustrated in Figure 5.4. These choices are derived in Appendix F.

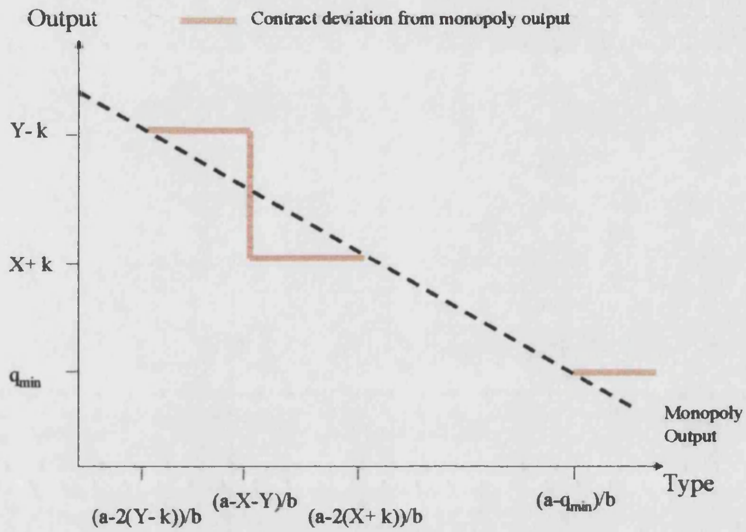

Figure 5.4: Output choices under Contract R^2

The semi-separating choices under the restricted two-interval contract are:

$$\begin{aligned} q &= q^M(\theta) \text{ if } \theta\epsilon \left[\theta_L, \frac{a-2(Y-k)}{b}\right] \\ q &= Y-k \text{ if } \theta\epsilon \left(\frac{a-2(Y-k)}{b}, \frac{a-(X+Y)}{b}\right] \\ q &= X+k \text{ if } \theta\epsilon \left(\frac{a-(X+Y)}{b}, \frac{a-2(X+k)}{b}\right] \\ q &= q^M(\theta) \text{ if } \theta\epsilon \left(\frac{a-2(X+k)}{b}, \frac{a-2q_{\min}}{b}\right] \\ q &= q_{\min} \text{ if } \theta\epsilon \left(\frac{a-2q_{\min}}{b}, \theta_H\right] \end{aligned}$$

5.3 Welfare comparison: contracts F, S and P

Table 5.2: Comparing the output floor contract to contracts S and P

Interval of Types	F	S	P
$\left[\theta_L, \frac{a-2q_{\min}}{b}\right]$	$q^M(\theta)$	$q^M(\theta)$	$q_{\min}$
$\left(\frac{a-2q_{\min}}{b}, \theta_H\right]$	$q_{\min}$	$q^M(\theta)$	$q_{\min}$

A comparison of the firm's output choices under the output floor contract (F) and the optimal fully separating (S) and optimal pooling (S) contracts is provided in Table 5.2. All types in the interval $\theta_L \leq \theta \leq \frac{a-2q_{\min}}{b}$ produce the same level of output under the output floor contract and the optimal fully separating contract. Those types who lie in the interval $\frac{a-2q_{\min}}{b} < \theta \leq \theta_H$ produce an output level higher than their monopoly level under the output floor contract. We therefore have a situation where, relative to the fully separating contract, no type chooses a lower level of output under the output floor contract and some types choose a higher level of output. As welfare is increasing in output for any given type we can conclude that the output floor contract increases welfare relative to the optimal fully separating contract. That is, *delegation with the output floor restriction is better than unrestricted delegation.*

We also find that the output floor contract yields a higher level of welfare than the optimal pooling contract. All types in the interval $\theta_L \leq \theta \leq \frac{a-2q_{\min}}{b}$ choose a higher level of output under the output floor contract than they produce under the optimal pooling contract. All other types produce the same level of output under both contracts. We therefore, again, have a situation where no type is choosing a lower level of output under the output floor contract and some types are choosing a higher level of output. Welfare is therefore higher under the output floor contract than under the optimal pooling contract. That is, *delegation with the output floor restriction is better than no delegation.*

The restriction on transfers means that the optimal fully separating and pooling contracts result in output levels which may be undesirably low. This problem can be at least partially overcome by delegating the decision to the firm and placing a lower-bound on the output choices[27]. This results in a semi-separating contract. This result contrasts with the standard literature on optimal regulatory mechanisms with unrestricted transfers which finds, for most standard belief functions, that the fully

[27] Armstrong and Vickers (2000) find a similar result in their analysis of multiproduct price regulation. They find that when demand is known, costs are unknown, and there is no lump-sum transfer, it is optimal to give the firm some discretion over its pricing policy (ie, give it discretion over individual tariffs while regulating overall revenues or a price index). They do not, however, explore the issue of what the optimal amount of discretion is.

separating contract is best[28].

5.4 Welfare comparison: contracts F and R^2

We assume here that the regulator chooses from a range of delegation contracts which place different degrees of restriction on the firm's choice set. The minimum level of restriction arises under the output floor contract, and the pooling contract offers the maximum level of restriction (ie, no choice). We have already shown that the minimum level is better than the maximum case of no delegation. The prevalence of output floor type contracts in practice might also suggest that the minimum level of restriction is better than any other restricted delegation contract. This has not, to our knowledge, been formally analysed. We therefore wish to examine the question of whether the output floor is in fact the best means of restricting the firm's choice set. We continue to assume that the regulated industry has an elastic demand curve and that the expected interval of costs satisfies the conditions described in section 5.1.4.

If it is the case that the output floor arrangement does not always yield the highest level of welfare, we also wish to identify the circumstances under which the regulator should place more restrictions on the firm's output choices. We examine this aspect of the regulation problem by comparing the output floor contract to the restricted two-interval contract. The question of whether even more restriction on the choice set would yield a higher level of welfare is explored in section 5.5, using a restricted m-interval contract (contract R^m, $m \geq 2$).

Table 5.3: Welfare Impact of Output Choices

Interval of Types	F	R^2	Welfare Change (given θ)
$I_0 = \left[\theta_L, \frac{a-2(Y-k)}{b}\right]$	$q^M(\theta)$	$q^M(\theta)$	0
$I_1 = \left[\frac{a-2(Y-k)}{b}, \frac{a-(X+Y)}{b}\right]$	$q^M(\theta)$	$Y-k$	+
$I_2 = \left[\frac{a-(X+Y)}{b}, \frac{a-2(X+k)}{b}\right]$	$q^M(\theta)$	$X+k$	-
$I_3 = \left[\frac{a-2(X+k)}{b}, \frac{a-2q_{\min}}{b}\right]$	$q^M(\theta)$	$q^M(\theta)$	0
$I_4 = \left[\frac{a-2q_{\min}}{b}, \theta_H\right]$	$q_{\min}$	$q_{\min}$	0

A high level comparison of the firm's output choices between the output floor contract and the restricted two-interval contract are provided in Table 5.3. We show the direction of the welfare change arising from the different output choices for each interval of types. For some types (I_1) there is an increase in welfare when we move

[28]We note that our results hold even if the regulator's belief function satisfies the monotone likelihood ratio property. This is because the optimal fully separating and optimal pooling contracts are independent of the belief function when the transfer function restriction is introduced.

from contract F to contract R^2, but for other types (I_2) there is a decrease in welfare. The net effect on expected welfare depends on the weight placed on each type interval, and on how the change in welfare from a change in output varies across types. These factors are explored in detail below.

5.4.1 Is Contract F always best?

In this section we consider the question of whether the output floor contract *always* yields the highest level of welfare from amongst the set of restricted delegation contracts. We find that there exists at least one case where an alternative contract, with more restriction than the output floor contract, yields a higher level of welfare. This result means that we cannot assume that the output floor contract is always best and a new dimension needs to be added to the design of regulatory delegation contracts which focuses on the optimal form of the choice set[29]. This analysis is of interest for industries with elastic demand and a relatively wide interval of potential cost types.

We use the following formulas to compare expected welfare under the alternative contracts.

The **level of expected welfare under contract** R^2 is:

$$E_\theta\left[W(q,\theta)\right] = \left\{ \begin{array}{c} \int_{\theta_L}^{\frac{a-2(Y-k)}{b}} W(q^M(\theta),\theta)f(\theta)d\theta + \int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} W(Y-k,\theta)f(\theta)d\theta \\ + \int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} W(X+k,\theta)f(\theta)d\theta + \int_{\frac{a-2(X+k)}{b}}^{\frac{a-2q_{\min}}{b}} W(q^M(\theta),\theta)f(\theta)d\theta \\ + \int_{\frac{a-2q_{\min}}{b}}^{\theta_H} W(q_{\min},\theta)f(\theta)d\theta \end{array} \right\} \tag{5.1}$$

With contract F expected welfare is as above but calculated at $k = \frac{Y-X}{2}$.

The **change in expected welfare** arising from a change in k is[30]:

$$\begin{aligned} \Delta &= \frac{\partial E_\theta\left[W(q,\theta)\right]}{\partial k} \\ &= \left\{ \int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta - \int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta \right\} \end{aligned}$$

The function Δ gives us the change in expected welfare arising from a small increase

[29] We use the phrase 'form of the choice set' to refer to the amount of restriction which is placed on the firm's choices. The question of where the restricted choice sets should lie on the output interval remains a central element of the design of the regulatory contracts, and becomes ever more complicated the more restrictions are introduced (ie, the more complex the form of the choice set). This element of the contract design is not considered in this paper.

[30] The calculation is based on the formula: $\frac{d}{dy}\int_{a(y)}^{b(y)} f(x,y)dx = \{f(b(y),y))b'(y) - f(a(y),y)a'(y)\} + \int_{a(y)}^{b(y)} \frac{df}{dy}(x,y)dx$

in k (ie, moving from Contract R^2 *towards* Contract F)[31]. As suggested by Table 5.3, only types in the regions I_1 and I_2 affect the change in welfare when the choice set is changed. If the firm's type is in I_1, it will choose a lower level of output when k is increased and there is a reduction in expected welfare. If the firm's type is in I_2, it will choose a higher level of output when k is increased and there is an increase in expected welfare.

If $\Delta \geq 0$ for all k we know that an increase in k will lead to an increase in welfare. That is, a choice set with a smaller set of disallowed output levels (less restriction) yields a higher level of welfare. Similarly, if $\Delta < 0$ for all k we know that an increase in k will lead to a decrease in welfare. That is, a choice set with a larger set of disallowed output levels (more restriction) yields a higher level of welfare.

Let $\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta$ **be called Integral X from this point on.**

Let $\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta$ **be called Integral Y from this point on.**

Δ will be positive if Integral X is greater than Integral Y, and it will be negative if Integral Y is greater than Integral X.

Theorem 5.1 *The output floor contract is not always best.*

Proof. (By example)

The regulator maximises a concave welfare function:

$$W = CS + \lambda\pi \quad \lambda\epsilon\,(0.5, 1)$$

The regulator does not know the firm's cost but has a belief that θ is distributed with a *non-increasing* density function, $f(\theta)$, $f'(\theta) \leq 0$.

The regulator delegates the output decision to the firm and considers offering it one of two alternative choice sets:

- Contract F: choose $q \in [q_{\min}, \infty]$
- Contract R^2 : choose $q \in ([q_{\min}, X + k] \cup [Y - k, \infty])$, $k = \left(\frac{Y-X-\varepsilon}{2}\right)$, ε small

The difference between these two choice sets is simply that contract R^2 introduces a small set of disallowed output levels, of size ε, in the output floor choice set. The two-interval choice set yields a higher level of welfare if a small increase in k from $\left(\frac{Y-X-\varepsilon}{2}\right)$ to $\left(\frac{Y-X}{2}\right)$ (the output floor case) results in a decrease in expected welfare.

[31]The reader is reminded that the parameter k determines the size of the interval of additional disallowed output levels, with $0 \leq k < \frac{Y-X}{2}$. *The smaller is k, the larger is the set of disallowed outputs.*

The impact of a small increase in k on expected welfare is equal to:

$$\Delta = \int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta - \int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta$$

Calculated at $k = \frac{Y-X-\varepsilon}{2}$ this is equal to:

$$\Delta\mid_{k_\varepsilon} = \int_{\frac{a-(X+Y)}{b}}^{\frac{a-(Y+X-\varepsilon)}{b}} \frac{\partial W(\frac{Y+X-\varepsilon}{2},\theta)}{\partial q(\theta)} f(\theta)d\theta - \int_{\frac{a-(Y+X+\varepsilon)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)} f(\theta)d\theta$$

A small increase from k to $\left(\frac{Y-X}{2}\right)$ leads to a reduction in welfare if:

$$\begin{array}{ccc} \left|\int_{\frac{a-(Y+X+\varepsilon)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)} f(\theta)d\theta\right| & > & \left|\int_{\frac{a-(X+Y)}{b}}^{\frac{a-(Y+X-\varepsilon)}{b}} \frac{\partial W(\frac{Y+X-\varepsilon}{2},\theta)}{\partial q(\theta)} f(\theta)d\theta\right| \\ \textit{Integral } Y & > & \textit{Integral } X \end{array}$$

We know that:

1. The intervals of integration are the same length.

2. Integral Y starts at a higher value than Integral X, for all $\lambda \epsilon (0.5, 1)$:

$$\begin{aligned} \frac{\partial W(\frac{Y+X+\varepsilon}{2}, \frac{a-(Y+X+\varepsilon)}{b})}{\partial q(\theta)} &= \frac{Y+X+\varepsilon}{2b} \\ \frac{\partial W(\frac{Y+X-\varepsilon}{2}, \frac{a-(X+Y)}{b})}{\partial q(\theta)} &= \frac{Y+X+(2\lambda-1)\varepsilon}{2b} \\ \frac{Y+X+\varepsilon}{2b} &> \frac{Y+X+(2\lambda-1)\varepsilon}{2b} \end{aligned}$$

3. With the *concave welfare function* (Lemma F.3):

$$\frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)} > \frac{\partial W(\frac{Y+X-\varepsilon}{2},\theta)}{\partial q(\theta)} \text{ for a given } \theta$$

4. The rate of change in welfare is decreasing in type for both output levels and the rate of decrease is the same (Lemma F.5):

$$\frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)\,\partial\theta} = \frac{\partial W(\frac{Y+X-\varepsilon}{2},\theta)}{\partial q(\theta)\,\partial\theta} = -\lambda$$

Points 3 and 4 imply that for $\theta \in I_1$ and $\theta' \in I_2$ $(\theta < \theta')$:

$$\frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)} > \frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta')}{\partial q(\theta)}$$

and:
$$\frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)} > \frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta')}{\partial q(\theta)} > \frac{\partial W(\frac{Y+X-\varepsilon}{2},\theta')}{\partial q(\theta)}$$

That is, the magnitude of the welfare decrease in the region I_1 is larger than the welfare gain in the region I_2. The impact on expected welfare also depends on the regulator's expectation about which region the firm's type lies in.

As $f(\theta)$ is non-increasing across the regions of interest (ie, $f(\theta) \geq f(\theta')$) we have:

$$\frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)} f(\theta) > \frac{\partial W(\frac{Y+X-\varepsilon}{2},\theta')}{\partial q(\theta)} f(\theta')$$

That is, the density function reinforces the direction of the relative magnitude of the change in welfare arising from a change in output.

As we are integrating these two functions over the same length interval, and the integral for types in the region I_1 starts at a higher point than the integral for types in the region I_2, we can conclude that the integrals retain the same relative magnitude as the integrands. This gives us, as required:

$$\left| \int_{\frac{a-(Y+X+\varepsilon)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(\frac{Y+X+\varepsilon}{2},\theta)}{\partial q(\theta)} f(\theta) d\theta \right| > \left| \int_{\frac{a-(X+Y)}{b}}^{\frac{a-(Y+X-\varepsilon)}{b}} \frac{\partial W(\frac{Y+X-\varepsilon}{2},\theta)}{\partial q(\theta)} f(\theta) d\theta \right|$$

That is, an increase in k to $\left(\frac{Y-X}{2}\right)$ leads to a reduction in welfare. This means that the two-interval restricted contract, with the additional interval of disallowed output choices, yields a higher level of welfare than the output floor contract.

We therefore conclude that the output floor contract is not a local optimum when the welfare function is concave and the regulator's belief function is non-increasing in type. This is because there exists at least one restricted two-interval contract, in the 'neighbourhood' of the output floor contract, which yields a higher level of welfare. There is thus at least one case in which the output floor contract is not best. ■

Figure 5.5 provides an extreme example of a density function which is non-increasing. There is a high density for types in the region I_1 and a very low (approximately zero) density for types in the region I_2. We therefore know that there is a much higher weight placed on the decrease in welfare in region I_1 relative to the increase in welfare in the region I_2, and hence expected welfare decreases if we move from contract R^2 to contract F. We also find, as derived in Appendix F, that if the *regulator's beliefs are uniform* the output floor contract is not best.

Theorem 5.1, and the examples of density functions which meet its conditions, suggest that, if the regulator places a lower weight (or the same weight) on the firm being inefficient relative to the weight placed on the firm being efficient, a restricted contract with two intervals will yield a higher level of welfare than the output floor

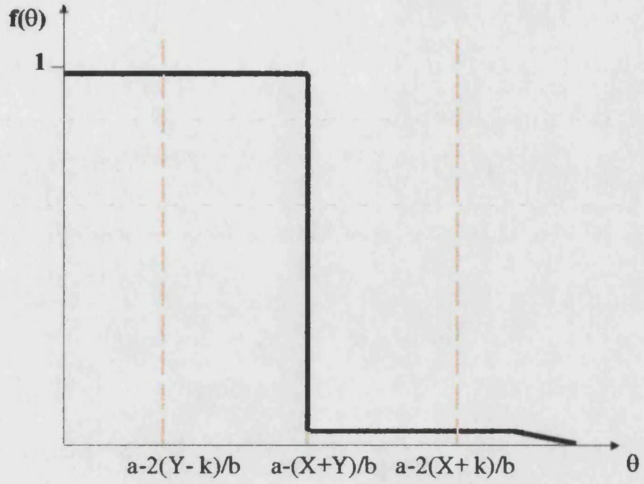

Figure 5.5: Example of the regulator's belief function

contract. This implies that, in industries with elastic demand and cost uncertainty (i.e. a relatively wide interval of potential cost types), simple output floor arrangements are not always best and, hence, **regulators should consider the possibility of introducing further restrictions on the firm's choice set when delegating output choices to them.**

5.4.2 Choosing between contract F and contract R^2

We know, from Theorem 5.1, that it is not always optimal for the regulator to offer the firm the output floor contract when delegating the output choice. The regulator must therefore consider a new dimension to the regulatory mechanism design problem: under what circumstances should the output floor contract be offered to the firm, and under what circumstances should additional restrictions be placed on this contract? We assume throughout that demand is elastic and that the interval of cost types is wide. We first present sufficient conditions under which the restricted two-interval contract yields a higher level of welfare than the output floor contract. We then find sufficient conditions under which the output floor contract yields a higher level of welfare than the restricted two-interval contract. These sufficient conditions are intended to provide a high-level guide as to the circumstances under which the regulator should choose one form of restricted delegation contract over another. In section 5.5 we use a simple simulated example to consider the case for extending the number of restrictions in the choice set further.

Dominance of contract R^2

We find that the regulator will be better off offering the firm contract R^2 rather than contract F if his beliefs are non-increasing and concave with type[32]. This result implies that if the regulator places a higher probability on the firm being efficient, relative to it being inefficient, then welfare will be increased if an output floor arrangement is replaced with a delegation contract where some of the choices under the output floor arrangement are now disallowed.

Lemma 5.1 *Suppose locally the regulator's belief function, $f(\theta)$, is non-increasing and concave. Then welfare from a restricted two-interval choice set increases with the size of the interval of disallowed output levels. That is, the more restriction there is in the two-interval choice set (ie, the lower is k) the higher is the level of welfare. If the welfare function is convex, the belief function only needs to be non-increasing for this to hold*[33].

Proof. See Appendix F ■

Proposition 5.1 *Suppose the conditions of lemma 5.1 hold. Then welfare is higher under contract R^2 than under contract F.*

Proof. The choice set under Contract F corresponds to a restricted 'two interval choice set' in which k is at its maximum value $\frac{Y-X}{2}$. We know, from Lemma 5.1, that any Contract R^2 which has a lower value of k will yield a higher level of welfare under the conditions of the Proposition. The regulator is therefore better off with Contract R^2 than Contract F. ■

We have shown that if the regulator's beliefs are non-increasing and concave in type then the more restricted delegation contract - Contract R^2 - is preferred to the output floor contract. Essentially this means that if the regulator places more weight on the firm being efficient rather than inefficient it will be optimal for a more restricted choice set to be included in the delegation contract. This implies that a firm which is a 'good' type, because it has low cost, is in some sense penalised because it is possible to induce it to choose a higher level of output with a more restricted contract.

[32]If the welfare function is convex (ie, $\lambda < \frac{1}{2}$) the belief function only needs to be non-increasing. We treat the weight on profits in the welfare function as exogenous here and hence assume that one case or the other (ie, concavity or convexity) applies. In practice, however, the regulator may choose the value of λ to reflect his belief about the appropriate distribution of rents between consumers and the firm. The endogeneity of this decision will need to be taken into account when determining the welfare-maximising form of the delegation contract. This is an aspect of the contract design problem which should be explored in future research.

[33]We know, from Appendix F.1, that the impact of the output level on the rate of change in welfare arising from a change in output depends on whether the welfare function is concave ($\lambda > \frac{1}{2}$) or convex ($\lambda < \frac{1}{2}$). The analysis of the impact of a change in the size of the prohibited interval in the decision set therefore varies with the value of λ and each possibility must be considered separately.

Dominance of contract F

We have found sufficient conditions under which the restricted two-interval contract yields a higher level of welfare than the output floor contract. In practice it is the latter contract which is observed, and this is also considered to be the simplest and hence most practical delegation arrangement to use. We therefore wish to ensure that the simple output floor contract is used whenever it yields a higher level of welfare than alternative restricted delegation contracts.

We determine sufficient conditions under which the output floor contract yields a higher level of welfare than a contract with a more restricted choice set. The conditions we find are, perhaps surprisingly given the prevalence of this form of contract in practice, more complex and less intuitive than those found for the restricted two-interval contract to be preferable. At a general level, however, the conditions suggest that if the regulator places a higher probability on the firm being inefficient rather than efficient, and if the extent of the difference is sufficiently large, then the output floor contract will yield a higher level of welfare than the alternative contract with extra restrictions on the choice set.

Taken together Propositions 5.1 and 5.2 provide a *high-level guide* to regulators on the conditions under which the output floor contract should and should not be used. As sufficient rather than necessary conditions the requirements of the Propositions may be overly demanding. It is a useful first step in the analysis of the optimal form of delegation contracts however.

Lemma 5.2 *Suppose locally:*

(i) the regulator's belief function, $f(\theta)$, is increasing and convex;

(ii) the rate of increase is large enough to offset the decreasing rate of change in welfare:

$$\frac{\partial f(\theta)}{\partial \theta} \geq \frac{Y}{X} > 1;$$

(iii) the degree of curvature of the belief function is greater than a positive integer:

$$\left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) > \frac{b}{X}; \textit{and}$$

(iv) the degree of convexity is sufficiently large:

$$\left(\frac{\partial^2 f(\theta)}{\partial \theta^2}\right)\left(\frac{1}{\frac{\partial f(\theta)}{\partial \theta}}\right) > \frac{2b\lambda}{X(1-2\lambda)} > 0.$$

Then welfare from a restricted two-interval contract decreases with the size of the interval of disallowed output levels. That is, the more restriction there is in the two-interval choice set (ie, the lower is k) the lower is the level of welfare. With a concave welfare function convexity of the belief function is not required for this to hold.

Proof. See Appendix F. ■

Proposition 5.2 *Suppose the conditions of Lemma 5.2 hold. Then contract F yields a higher level of welfare than contract R^2.*

Proof. The choice set under Contract F is essentially a 'two interval choice set' in which k is at its maximum value $\frac{Y-X}{2}$. We know, from Lemma 5.2, that any Contract R^2 which has a lower value of k will yield a lower level of welfare under the conditions of the Proposition. The regulator is therefore better off with Contract F than Contract R^2. ■

We have shown that if the regulator's beliefs are sufficiently increasing and sufficiently convex, Contract F is preferable to a contract which places an additional restriction on the firm's choice set. We note that more conditions are required on the density function here compared to the sufficient conditions needed for contract R^2 to yield a higher level of welfare. It is also interesting that the sufficient conditions are dependent on the parameter values X and Y in this situation. In particular, the size of the lower boundary, X, and the distance between the boundaries (ie, the ratio of X to Y) is a key influence on whether or not these sufficient conditions are met. The conditions are not immediately intuitive but they suggest that the regulator must place a sufficiently higher weight on the firm being inefficient rather than efficient for expected welfare to be higher with the output floor contract.

We also find, as shown in Corollary 5.1, that if the conditions are met so that the output floor contract yields a higher level of welfare than Contract R^2, then Contract F will yield a higher level of welfare than any other delegation contract incorporating more restrictions on the firm's choice set.

Corollary 5.1 *Suppose the conditions of Lemma 5.2 hold. Then contract F will yield a higher level of welfare than any restricted m-interval contract, contract R^m, which has $m > 1$ intervals of disallowed output levels.*

Proof. (Sketch)

Assume the conditions of Proposition 5.2 hold. Take two contracts. The first, contract R^{m-1}, has m-1 sets of disallowed output levels. The second contract, R^m, is equivalent to contract R^{m-1}, except that an additional set of disallowed output levels has been introduced. By the same logic as the proofs of Lemma 5.2 and Proposition

5.2, the regulator prefers the contract with $m-1$ intervals to the contract with m intervals.

Expected welfare is higher when the gap between two intervals is closed, independent of the position of that gap. The Proposition can therefore apply to the comparison of any two choice sets, where one is equivalent to the other except for the fact that, in one section of the choice set, two of the intervals of the contract with m intervals are combined to form a contract with $m-1$ intervals.

By iteration we can see that a contract with $m-2$ intervals is preferred to a related contract with $m-1$ intervals, a contract with $m-3$ intervals is preferred to a related contract with $m-2$ intervals, and so forth. We therefore have a hierarchy of contracts, where R^j is a choice set with j intervals, such that:

$$F \succ R^2 \succ R^3 \ldots\ldots\ldots\ldots \succ R^{m-1} \succ R^m$$

We conclude that if the conditions hold for the output floor contract to be preferred to the restricted two-interval contract, then the output floor contract is preferred to all delegation contracts which place *additional* restrictions on the firm's choice set. ■

5.5 Introducing further restrictions

We have found that, given the assumptions of our model, the output floor contract yields a higher level of welfare than unrestricted delegation (the fully separating contract) and no delegation (the pooling contract). This suggests that some form of restricted delegation contract is best. When we focus on a set of restricted delegation contracts we find that the output floor contract, despite its prevalence in practice, is not always best. The characteristics of the regulator's belief function determines whether the introduction of one additional restriction on the firm's output choice set enhances welfare relative to the output floor contract.

We assume that the industry characteristics discussed in section 5.1.4 and the conditions of Lemma 5.1 hold here, and extend the analysis to determine whether increased levels of restriction lead to higher welfare levels. We only consider contracts which include a countable number of intervals. Extending the mathematical analysis to formally compare contract F to a restricted m-interval contract is overly complex, and adds little to our general result that, in some situations, more restriction is better than less. We therefore use a simulated example to examine the case for introducing more disallowed output choice sets in the restricted delegation contract[34].

[34]Simulations are used in several seminal papers on the literature comparing welfare under a range of regulatory contracts. For example, Gasmi, Ivaldi and Laffont (1994) and Schmalensee (1989) base their analysis of various incentive schemes on the results of a detailed simulation exercise. Armstrong, Rees and Vickers (1995) also use simulations when analysing the optimal length of the regulatory lag.

When considering restricted m-interval contracts we expect that, under the same conditions as are found in Lemma 5.1, the regulator prefers to have more intervals in the choice set rather than less. We stress that a contract with m intervals is an adjusted form of the contract with m-1 intervals, where an extra interval of disallowed choices has been added. That is, we are not considering arbitrarily chosen contracts with different numbers of intervals, but are focusing on the impact of amending an existing contract by introducing an extra interval of disallowed output levels in the delegated choice set.

In our simulated example we assume that the regulator's belief function has a uniform distribution. In this situation we know that the restricted two-interval contract yields a higher level of welfare than the output floor contract. We wish to confirm that as we increase the number of intervals further there is an on-going increase in welfare. From a practical perspective we also want to determine the marginal welfare gain from an extra interval, as the regulator will presumably wish to trade-off the marginal cost of increased complexity in the delegated contract against the marginal benefit.

The set-up for our example is as follows.

- There is a simple linear demand curve, $q = 100 - p$.
- The firm's marginal cost lies in the interval $[\theta_L, \theta_H] = [5, 95]$. With this we know that the only feasible output values lie in the region $(0, 95]$ and the monopoly output levels lie in the region $\left[q^M(\theta_H), q^M(\theta_L)\right] = [2.5, 47.5]$. The interval is sufficiently wide to meet the industry characteristics discussed in section 5.1.4.
- The regulator believes that the firm's marginal cost has a uniform distribution: $\theta \sim U$ on the interval $[\theta_L, \theta_H]$, $f(\theta) = \frac{1}{\theta_H - \theta_L}$.
- The regulator delegates the final output decision to the firm but restricts the firm's choice set. We discuss below the range of restrictions which the regulator considers.
- The firm chooses the level of output from the restricted choice set which maximises profit.
- The regulator chooses the regulatory contract which, given the firm's expected choices, yields the highest level of welfare: $W = CS + \lambda\pi$. We consider the cases of $\lambda = \{0.25, 0.5, 0.75\}$.

We consider five alternative delegation contracts. Appendix F explains how the simulations for each contract option were carried out.

Case1: Unconstrained - the regulator delegates the output choice to the firm and sets no restrictions on the firm's choices. This results in each type choosing its monopoly output level and corresponds, as we have shown in Section 5.2, to the outcome from the optimal fully separating contract.

Case 2: Output floor - the regulator delegates the output choice to the firm and sets an output floor on the firm's choices. We have shown, in Section 5.2, that the optimal floor corresponds to the first-best output level of the inefficient type. This gives us: $q_{\min} = 5$. Demand is elastic at this output floor.

Case 3: Two intervals - the regulator delegates the output choice to the firm, sets the same output floor as in Case 2, and adds an extra restriction that outputs between two points, X and Y, cannot be chosen by the firm. We find the values of X and Y that yield the highest level of welfare.

Case 4: Three intervals - the regulator delegates the output choice to the firm, takes the welfare-maximising contract in Case 3, and adds an extra restriction that outputs between points W and Z cannot be chosen by the firm. Two options are considered.

1. choose W and Z to the right of X and Y; and
2. choose W and Z to the left of X and Y.

From amongst these options we find the values for W and Z which maximise welfare.

Case 5: Four intervals - the regulator delegates the output choice to the firm, takes the welfare-maximising contract in Case 4, and adds an extra restriction that outputs between points U and V cannot be chosen by the firm. Three alternative options are considered.

1. choose U and V to the right of {W, Z, X, Y}[35];
2. choose U and V in the middle of {W, Z, X, Y};
3. choose U and V to the left of {W, Z, X, Y}.

From amongst these options we find the values for U and V which maximise welfare.

An example of the contract options for cases 3, 4 and 5 are illustrated in Figure 5.6. Tables G.1, G.2 and G.3 in Appendix G show the interval values considered, and the corresponding welfare levels. The main welfare comparisons are discussed here.

[35]The Case 4 interval points may alternatively be ordered as {X, Y,W, Z}.

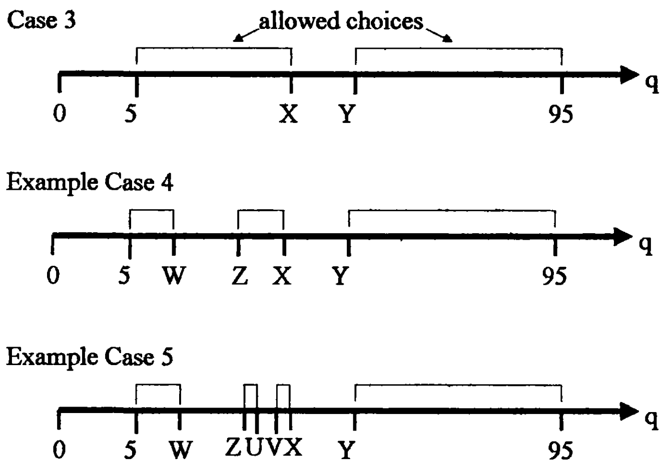

Figure 5.6: Example simulated contract options

Figures 5.7, 5.8 and 5.9 show the firm's output choices with the two, three and four interval contracts when $\lambda = 0.25$ [36]. The allowed output intervals are as follows:

$$\begin{aligned} \text{Two} &: \quad [5, 29.88] \cup [36.62, 47.5] \\ \text{Three} &: \quad [5, 9.79] \cup [29.74, 29.88] \cup [36.62, 47.5] \\ \text{Four} &: \quad [5, 9.79] \cup [29.74, 29.75] \cup [29.78, 29.88] \cup [36.62, 47.5] \end{aligned}$$

The output choices follow the pattern expected. As the number of intervals increases, the impact of the additional interval is reduced when it is very close to an existing interval and/or is very narrow. This is what happens with the extra interval in the four interval contract.

[36]The values are slighly distorted as the simulations restricted the output choices to be in intervals of 0.25 which in some cases meant that very small intervals (ie, with bounds which were less than 0.25 apart) had no effect on output choices. The impact on welfare is expected to be minimal, however, as such tight intervals would not alter the output choices, and hence welfare, significantly.

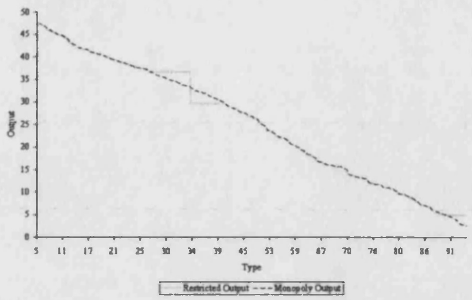

Figure 5.7: **Two interval contract output choices**

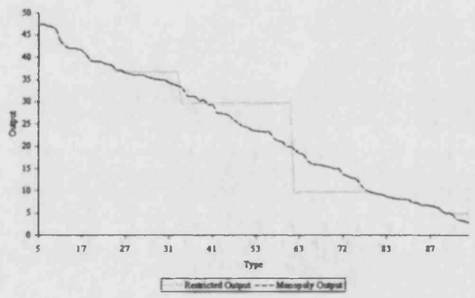

Figure 5.8: **Three interval contract output choices**

Figure 5.10 shows the level of welfare from each of the contract choices for the three values of λ. We see that welfare increases with the number of intervals for all values of λ. Welfare, not surprisingly, varies with the value of λ, and the marginal change in welfare from an increase in the number of intervals also varies with the value of λ.

Table 5.4 shows the *marginal* welfare change from increasing the number of intervals. In the case of a concave welfare function ($\lambda = 0.5, 0.75$) the size of the percentage change increases when we go from none to one, one to two, and two to three intervals. It then declines, however, when we increase the number of intervals to four. For the convex welfare function ($\lambda = 0.25$) there is always an increasing marginal gain from increasing the number of intervals.

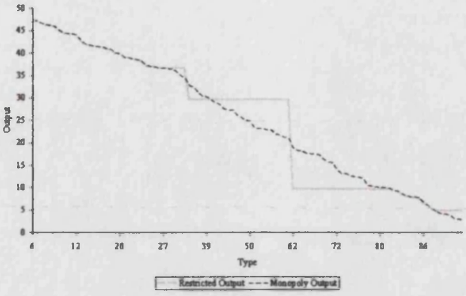

Figure 5.9: **Four interval contract output choices**

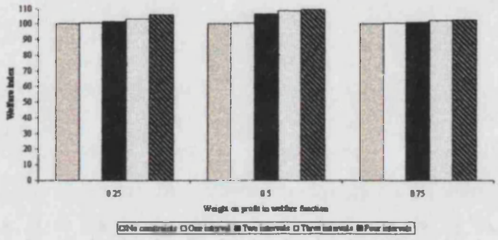

Figure 5.10: Welfare under restricted delegation contracts

Table 5.4: Marginal change in welfare

Interval Change	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 0.75$
None to one	2.67	3.30	3.43
%	0.45	0.42	0.35
One to two	5.40	48.37	6.44
%	0.91	6.08	0.65
Two to three	11.94	15.81	11.69
%	1.99	1.87	1.17
Three to four	17.97	6.05	5.10
%	2.93	0.70	0.50

We find, in our simulated example, that welfare increases with the level of restriction placed on the choice set. Further research is needed to determine where the regulator stops when considering how many discrete intervals of disallowed choice sets to include in the contract. It is not clear at this stage what the optimal restricted contract looks like in the limit. We expect that, in a practical setting, there will be situations where the welfare gain from more intervals is traded-off against the cost of designing complex delegation contracts.

While we cannot estimate the actual costs which would arise in the regulatory office, we can give an indication of the extra complexity that arises when we increase the number of intervals in the contract. When running the simulations for the two interval contract we had to find the values of X and Y which maximised welfare for the given value of λ . When we moved to three intervals, we had to consider two scenarios when finding W and Z for a given value of λ - to the left and to the right of the X and Y from the optimal two interval contract. Our workload (or at least that of the Matlab© programme) increased twofold. This was reflected in the time taken to determine the three interval contract which yielded the highest level of welfare. Moving to the four interval case increased the complexity of finding the optimal contract even further. We had three scenarios to consider here, and the workload was increased relative to the three interval case and relative to the two interval case. There is thus an increasing cost to increasing the number of restrictions in the delegation contract. This cost needs to be considered alongside the marginal welfare gain discussed above.

This example suggests that when the belief function is uniform, an increase in the number of restricted choice intervals in the delegation contract increases welfare. We have also shown that there is an increased amount of effort required to generate more complicated contracts. Further analysis is required to determine if this is true for other distributions which satisfy the conditions of Proposition 5.1.

5.6 Conclusions

Fixed price cap contracts are amongst the most prevalent regulatory mechanisms used in practice[37]. These contracts provide strong incentives for the regulated firm to improve technical efficiency but they do not yield maximum welfare levels. In particular there is a risk, in industries with significant cost uncertainty (i.e. with wide intervals of potential cost types) and elastic demand, that the regulated firm will continue to price at the monopoly level under the price cap contract. It is therefore appropriate to consider the case for introducing alternative forms of fixed price contract which may yield higher levels of welfare[38]. We note however that the problem is unlikely to be significant in the regulated water and electricity sectors in the UK as demand is inelastic in these industries. The analysis is therefore of more relevance for other sectors which face elastic demand functions.

With this in mind we compared welfare under the standard price cap contract to a number of different contracts which delegate the price choice to the firm but

[37]We move to an analysis of price cap regulation in our concluding remarks as this is a more frequent mode of regulation than output floors in practice. We stress, again, that all conclusions reached in this paper for output floors automatically translate to the mirror-case of price caps.

[38]It is expected that the fixed nature of the contract will ensure that technical efficiency incentives are retained even if the form of the contract is changed. The strength of the incentives may change, however. This aspect of the contract design problem warrants further attention in future research.

place more, or less, restriction on the firm's allowed choice set. The level of restriction determines the *form* of the contract. The comparisons considered were a no delegation contract (the pooling case), an unrestricted delegation contract (the fully separating case) and a restricted two-interval contract which incorporates the choice set under the price cap contract but with an additional set of disallowed price levels.

We found, using a single-period adverse selection model with restricted transfers, that the price cap contract yields a higher level of welfare than the no delegation contract and the unrestricted delegation contract. This implies that some form of restricted delegation contract is best, but it is not clear that the cap is necessarily the welfare-maximising level of restriction to place on the firm's choice set. Indeed an analysis of alternative restricted delegation contracts showed that the price cap contract is not always best. This result introduces an extra dimension to the design of regulatory contracts. The regulator first needs to determine what the optimal level of restriction is, and then determines the welfare-maximising boundaries in that contract. We have not considered the second element of the regulator's problem here.

We found sufficient conditions under which the price cap contract is best, and sufficient conditions under which a two-interval contract with an additional restriction to the cap is best. In the latter case we also found, using a simulated example, that increasing the level of restriction further - with three- and four-interval contracts - yields increasing levels of welfare. We noted, however, that the cost of creating overly complex contracts would need to be traded-off against this welfare gain in a practical setting. Further research is needed to determine what the optimal number of such intervals is.

The sufficient conditions under which the price cap contract is and is not best relate to the characteristics of the regulator's belief function; in particular, whether or not it is increasing, and whether it is convex or concave. Many practitioners criticise the mechanism design literature on the grounds that any contract which requires the distribution function of the firm's type to be common knowledge cannot be used in practice[39]. We argue, however, that the sufficient conditions of our propositions can be used in a general way in practice, without the regulator needing to know the full density function. For example, if the regulator has a very high belief that the firm is efficient we can conclude, along the lines of the proof to Proposition 1, that the price cap contract is not best. Similarly, if the regulator has a very strong belief that the firm is inefficient - skewing the density function towards the high types - we conclude that the price cap contract is best. We recognise, however, that as sufficient but not necessary conditions our results are of limited use to regulators if they have beliefs which do not fall neatly into the conditions derived.

[39]See, for example, Crew and Kleindorfer (2002).

This paper contributes to the literature on optimal regulatory mechanisms in three key ways.

1. We presented a formal proof of how the restricted transfer function affects the standard mechanism design problem. This enabled us to determine the optimal fully separating and pooling mechanisms, and allowed us to compare their welfare properties to those of the price cap contract. It is important that these welfare comparisons are made explicit rather than being presumed, particularly when less restrictive constraints on the transfer function may allow for pooling, fully separating or semi-separating mechanisms to yield the highest level of expected welfare.

2. In contrast to the theory and practice of price cap regulation, and the literature on optimal delegation in a single period, we have not presumed that the regulator can only restrict the firm's choice set to be a continuous interval. In this way we extend the framework for designing a regulatory contract to explicitly include an analysis of the level of restriction, or form, of the contract.

3. We have shown that the price cap contract is not always best, and we found conditions under which more restriction on the firm's choice set is better than less. From a policy perspective this suggests that regulators in sectors with elastic demand and cost uncertainty should, in some circumstances, consider amending the form of the price contracts used in practice.

Our comparison of the restricted two-interval contract to the price cap contract is based on a limited model, and a number of extensions could be considered in future research to better reflect actual regulatory environments. First, we could allow the firm to reduce its costs by undertaking effort. At a high level, we would expect that if the regulatory lag and other parameters are the same in the fixed price cap contract and the fixed two-interval contract, then the firm would have an incentive to reduce costs in each case as it would retain the profits earned for itself[40]. The complication in the two-interval case is that the effort choice will, by reducing costs, affect the profit-maximising price level. In addition, the output level will affect the firm's effort choice as it determines the marginal benefit of each unit of effort. Given this endogeneity between the price choice and the effort choice, the firm may need to make trade-offs between the optimal level of effort and the possibility that the new cost level will push the optimal price into a disallowed interval. This may encourage the firm to undertake

[40]Laffont and Tirole (1993, Ch2) similarly suggest that, when the transfer is restricted, the 'main incentive issues' addressed in their analysis of optimal regulatory mechanisms do not change but the 'pricing issues' do.

less effort so that the profit-maximising price, given cost, is at a higher allowed level. Alternatively, it may encourage the firm to undertake more effort, so that a lower allowed price yields a higher level of profit. In general the incentive to reduce costs will be retained in any restricted m-interval contract if the contract is fixed for a set period. The extent of the incentives may be affected by the impact of the effort choice on feasible optimal price choices however.

A second extension involves the comparison of the contracts in a dynamic framework[41]. Again the impact on effort choices over time would need to be considered. The issues are likely to be similar to the static model with effort if the rules about sharing past savings, and using historical information to set the contract, are the same across all contract forms. We expect the firm to care about its reputation in the dynamic model and, as the price choice could reveal the firm's type, more bunching may occur at the extremes of any choice set.

We also need to consider the fact that the regulator's choice of contract may change over time. Consider the case of a regulator who has a very high belief that the firm is inefficient at privatisation. The optimal choice set is the price cap contract. The firm operates under this fixed-price contract and becomes more efficient. The regulator's beliefs change and, in time, we may reach a point where there is now a high belief that the firm is efficient. Under these circumstances a move to a choice set with more restriction is optimal. The impact of the incentive mechanism on the firm's cost level and the regulator's belief function may therefore result in the optimal form of the delegation contract changing over time.

A further extension of the dynamic analysis would involve the regulator being unable to commit to always using a restricted delegation contract and he may wish to move to the optimal fully separating or optimal pooling contracts as more information on the firm's costs becomes available. The absence of commitment to the contract form would, no doubt, result in a ratchet effect problem, and it would be useful to determine the impact of such 'mechanism cycles' on welfare.

[41] Legros (1993) analyses the impact of repetition on delegation contracts in a more general setting. He emphasises that the contract choice in the first period will be used as an information signal in future periods. This paper may provide a useful basis for examining welfare under our range of regulatory contracts in a dynamic framework.

Chapter 6

Conclusion

We have analysed the efficiency properties of the RPI-X regime in this thesis. We found that technical efficiency incentives have been delivered, but are not as strong as expected. This is demonstrated by the fact that there has been a significant decrease in real unit operating costs since privatisation, but only a moderate improvement in total factor productivity growth. Allocative efficiency is not achieved with this regulatory regime, and the cost-saving incentives have the potential to distort the firm's decision on output delivery. In addition, the regulators do not use willingness to pay information directly to set output targets, and hence there is concern that the quality of output produced is not efficient.

These properties are affected by the characteristics of the game played by the regulator and the firm, and by the way in which the cap itself is set. Bennett and Waddams Price (2002) reach a similar conclusion when they argue that 'Incentive mechanisms within utilities are limited both by the nature of the industries and the need to achieve political consensus in the design and outcome of their regulation'. In particular, the level of welfare delivered is dependent on the trade-offs which are made between allocative efficiency, technical efficiency, output delivery and distribution[1]. This vindicates our argument that a detailed understanding of the way in which a regulatory mechanism operates in practice is required if an accurate assessment of its welfare properties is to be made.

Given that welfare is not maximised with this regime, we questioned whether alternative mechanisms could be designed which might increase welfare[2]. The charac-

[1]This was noted, early on, by Helm and Yarrow (1988) who found that 'measures to improve allocative efficiency are frequently in substantial conflict with measures to promote cost efficiency, and interventions in respect of one aspect of company decision-making tend to produce perverse incentives in respect of other dimensions of choice'.

[2]According to Helm (1995a) the popular position on RPI-X regulation is that 'it is better than what we had before and better than the American alternative, so why change?'. Our detailed review of the welfare properties of the current RPI-X regime explains, we hope, why change is needed.

teristics of the game are assumed fixed by the legislative and institutional framework, and we focused on the design of the contract itself[3],[4]. We found, in the first instance, that technical efficiency might be increased by retaining the RPI-X mechanism but changing the way in which savings are shared with consumers. Similarly, quality of service might be improved, with the RPI-X contract remaining in place, if the choice between a range of price-quality contacts is given to consumer groups. We also found that the regulator might be able, in some instances, to increase welfare by placing more restriction than the cap on the firm's price choice set. This opens up a whole new dimension to the regulatory debate which focuses on the issue of what form of fixed price contract should be used. The feasibility of these potential improvements is limited to some degree by the nature of the regulated sector (e.g. whether or not demand is stochastic, whether or not demand is elastic, and whether or not there is significant cost uncertainty) and the regulator's asymmetric information set.

Sappington (2000) argues that all regulatory mechanisms have their advantages and disadvantages and that 'no regulatory plan is a perfect substitute for the discipline of competitive markets'. We are therefore not surprised to find that there are benefits and costs associated with the RPI-X regime, and that welfare is not maximised. This does not mean, however, that we should assume that the mechanism is as good as we can get in the absence of a competitive market. Efforts should be made to make the regulatory scheme better, even if it can never be first-best. We recognise, however, that there may be resistance to change.

Cox and Isaac (1987) argued that the preference to stick with rate-of-return regulation at that time was 'due more to historical precedence than to inherent superiority'. A similar argument could also be applied to those who claim that there is no need to change the current RPI-X mechanism. Inertia and tradition are not, however, sufficient reasons to ignore the possibility of increasing welfare within the regulatory regime.

Some of our proposals, particularly the use of contracts with more restriction on the firm's choice, would be difficult to implement. Complexity should also not be used as an excuse for staying with the current arrangement. Indeed, as noted by Vickers and Yarrow (1988a), 'Effective regulation is necessarily a complex business, and to

[3] A key lesson from our analysis is that the regulatory framework and regulatory mechanisms should be designed simultaneously, as the former affects the welfare properties of the latter. If we were devising an entirely new regulatory regime we would consider using a different framework - to change the characteristics of the game - as well as a different incentive mechanism. We assume, however, that we are constrained by the existing framework, as established in the privisation legislation for each sector, and we have focused out attention on changes which could be made within this framework.

[4] We stress, again, that it has only been feasible to consider a small sample of potential improvements in this analysis. Further ideas, including proposals to develop penalties for overforecasting, and an analysis of the case for franchising the right to operate the network business, should be explored in future research.

pretend otherwise is likely to have damaging long-term consequences for the industries concerned'. We therefore suggest that regulators consider how these proposals might be converted into feasible regulatory mechanisms given the specific characteristics of the firm's operating environment, rather than deciding in advance that they are too costly or difficult to introduce.

Our research has highlighted a number of areas to be explored further. The following extensions would aid our understanding of how the RPI-X mechanism affects welfare, and enable us to determine feasible ways of increasing welfare in the existing regulatory regime.

- Sappington (2000) emphasised that 'current estimates of the impact of incentive regulation may not reflect accurately its true long term impact'. This suggests that ongoing analysis of the welfare properties of the mechanism should be undertaken, particularly as the specific details underlying the regime change. This analysis should incorporate both theoretical arguments and empirical testing.

- A richer description of the RPI-X game could be developed which incorporates the consultation phase of the periodic review process. This would allow us to analyse the precise impact of the different interest groups, including the industry, on the regulator's decision-making process. It would also enable us to analyse in more detail the information transmission game which the firm and the regulator play when business plans are submitted.

- We have assumed that the firm makes decisions as a profit-maximiser with a ten-year horizon. In practice, managers and shareholders within the firm are likely to have different objectives, from each other and relative to profit-maximisation. The impact of alternative objectives, and hence choices, on the mechanism's efficiency properties should be explored further. This will help us to understand how future improvements to the mechanism should be designed so that they are appropriately focused on the party making the relevant decisions.

- Our analysis of the welfare impact of the RPI-X regime could be used as the basis for an empirical examination of the costs and benefits of this form of regulation. The research could be based on a similar methodology to that used by Newbery and Pollitt (1997) and Domah and Pollitt (2001). Such an exercise would, no doubt, be very difficult to undertake, particularly as there is no clear benchmark against which to compare the changes in prices, profits and total factor productivity. One possible counterfactual to consider is what would happen with a pure price cap. Any analysis of the welfare impact of the regime should include the costs of regulation itself. These costs have not been discussed here but are considered to be growing in all sectors[5].

- It would be interesting to undertake a similar detailed analysis of price cap regulation in the US, and to compare the welfare effects of these schemes to the RPI-X mechanism. The comparison should reflect both the way in which the price caps are designed and implemented, and the underlying characteristics of the games which are played between the firms and the regulators. This would enable us to determine the factors which have most influence on the level of welfare. The comparative analysis could be extended to price cap regimes in other countries. The Chilean electricity distribution industry might be of particular interest here as it has been in place since 1982 and it is, according to Di Tella and Dyck (2002), 'closer to many theoretical properties of pure price cap than other regimes'. The price cap schemes used in Australia would also be of interest as they were, at least originally, based on the UK regime[6].

- We have presented a selection of high-level changes which could be made to the regulatory contract to potentially increase welfare. Four of these focused on changes to the existing price cap contract, while the fifth was based on altering the form of the contract so that more restriction is placed on the firm's price choice. Further analysis of the impact of these proposals on total welfare is required. This should focus on the net effect of trade-offs between allocative and technical efficiency. In addition, an analysis of how these proposals could be translated into feasible mechanisms should be undertaken, taking account of information limitations and uncertainty about the firm's future operating environment.

- Finally, there are no doubt several other changes which could be made to the way in which the X-factor is set, and to the regulatory contract more generally. It is hoped that our description of the RPI-X game provides a framework in which the impact of such changes can be explored. Several authors have argued that the regulatory institutions and legal rules should be changed so as to increase welfare. These issues were discussed in detail in the government's 1997 review of utility regulation. We argue that it is also important to consider the case for improving the detail of the regulatory contract itself, within the confines of the existing legal and institutional structures. There are thus two related, and complementary, strands of research which require ongoing consideration.

[5]See the study by WS Atkins and OXERA (2001) for further information on the increased running costs of the regulators' offices.

[6]Forsyth (1999) presents a description and critique of the price cap schemes used in Australia.

Bibliography

[1] Ai C and Sappington D (2002), 'The Impact of State Incentive Regulation on the US Telecommunications Industry', *Journal of Regulatory Economics,* Vol 22, pp133-160

[2] Akmal M and Stern DI (2001), 'The structure of Australian residential energy demand', *Working Papers in Ecological Economics, The Australian National University Centre for Resource and Environmental Studies,* Number 0101.

[3] Anderson T and Hsiao C (1982), 'Formulation and estimation of dynamic models using panel data', *Journal of Econometrics,* Vol 18, pp47-82

[4] Arellano M (1989), 'A note on the Anderson-Hsiao estimator for panel data', *Economics Letters,* Vol 31, pp337-341

[5] Armstrong M (2003), 'Privatisation, regulation and competition', *Paper prepared for a presentation at the conference "Competition and Development",* hosted by the Competition Commission and the Competition Tribunal of South Africa

[6] Armstrong M (1999), 'Optimal regulation with unknown demand and cost functions', *Journal of Economic Theory,* Vol 84, pp196-215

[7] Armstrong M (1994), *Delegation and Discretion,* Discussion Papers in Economics and Econometrics, University of Southampton, Discussion Paper No 9421

[8] Armstrong M, Cowan S and Vickers J (1994), *Regulatory Reform: Economic Analysis and British Experience,* The MIT Press

[9] Armstrong M, Rees R and Vickers J (1995), 'Optimal regulatory lag under price cap regulation', *Revista Espanola de Economia,* Vol 12, pp93-116

[10] Armstrong M and Sappington D (2002), 'Recent developments in the Theory of Regulation', in *Handbook of Industrial Organisation (Vol III),* forthcoming

[11] Armstrong M and Vickers J (2000), 'Multiproduct price regulation under asymmetric information', *The Journal of Industrial Economics*, Vol 48, pp137-160

[12] Armstrong M and Vickers J (1995), 'Competition and regulation in telecommunications', in Bishop M, Kay J and Mayer C (eds), *The Regulatory Challenge*, Oxford University Press

[13] Arocena P and Waddams Price C (2001), 'Generating Efficiency: Economic and Environmental Regulation of Public and Private Electricity Generators in Spain', *International Journal of Industrial Organization*, Vol 20, pp41-69

[14] Athey S and Roberts J (2001), 'Organizational design: decision rights and incentive contracts', *The American Economic Review*, Vol 91, pp200-205

[15] Averch H and Johnson L (1962), 'Behaviour of the firm under regulatory constraint', *The American Economic Review*, Vol 52, pp1052-1069

[16] Bagnoli M and Bergstrom T (1989), 'Log-concave probability and its applications', *CREST WP #89-23*, University of Michigan

[17] Baltagi B (1995), *Econometric Analysis of Panel Data*, Wiley

[18] Banerjee A and Dasgupta K, 'Does Incentive Regulation "Cause" degradation of retail telephone service quality?', *Presentation at the Rutgers University Advanced Workshop in Competition*, 20th Annual Conference

[19] Baron (1989), 'Design of regulatory mechanisms and institutions', in Schmalensee R and Willig R (eds), *Handbook of Industrial Organization, Volume II*, Elsevier Science Publishers, pp1347-1447

[20] Baron D and Besanko D (1987), 'Commitment and fairness in a dynamic regulatory relationship', *The Review of Economic Studies*, Vol 54, pp413-436

[21] Baron D and Besanko D (1984), 'Regulation, asymmetric information and auditing', *RAND Journal of Economics*, Vol 15, pp447-470

[22] Baron DP and Myerson RB (1982), 'Regulating a monopolist with unknown costs', *Econometrica*, Vol 50, pp911-930

[23] Baumol W (1995), 'Modified regulation of telecommunications and the public-interest standard', in Bishop M, Kay J and Mayer C (eds), *The Regulatory Challenge*, Oxford University Press

[24] Beesley ME (1997), *Privatization, Regulation and Deregulation*, Routledge, 2nd edition

[25] Beesley ME (1996a), 'RPI-X principles and their application to gas', *Regulating Utilities: A Time for Change*, The Institute of Economic Affairs. Reproduced in Beesley ME (1997), *Privatization, Regulation and Deregulation*, Routledge, 2nd edition

[26] Beesley ME (1996b), 'The conditions for effective utility regulation', Presentation to the Australian Competition and Consumer Commission, February. Reproduced in Beesley ME (1997), *Privatization, Regulation and Deregulation*, Routledge, 2nd edition

[27] Beesley ME (1994), 'RPI-X principles and their applications to gas', Lecture from *IEA Readings 44: Regulating Utilities: A Time for Change*. Reproduced in Beesley ME (1997), *Privatization, Regulation and Deregulation*, Routledge, 2nd edition

[28] Beesley MC and Littlechild SC (1989), 'The regulation of privatized monopolies in the United Kingdom', *RAND Journal of Economics*, Vol 20. Reproduced in Beesley ME (1997), *Privatization, Regulation and Deregulation*, Routledge, 2nd edition

[29] Beesley ME and Littlechild SC (1983), 'Privatization: principles, problems and priorities', *Lloyds Bank Review*. Reproduced in Beesley ME (1997), *Privatization, Regulation and Deregulation*, Routledge, 2nd edition

[30] Bennett M and Waddams Price C (2002),'Incentive contracts in utility regulation', in Brousseau E and Glachant J (eds), *The Economics of Contracts: Theories and applications*, Cambridge University Press

[31] Bernstein JI and Sappington D (1999), 'Setting the X Factor in Price-Cap Regulation Plans', *Journal of Regulatory Economics*, Vol 16, pp5-25

[32] Besanko D and Spulber D (1992), 'Sequential-equilibrium investment by regulated firms', *RAND Journal of Economics*, Vol 23, pp153-170

[33] Bishop M, Kay J and Mayer C (1994), 'Introduction', *Privatization and Economic Performance*, Oxford University Press

[34] Blundell and Bond (1999), *GMM Estimation with Persistent Panel Data: An Application to Production Functions*, The Institute for Fiscal Studies, Working Paper Series No W99/4

[35] Bolton P and Dewatripont M (2002), *Contract theory,* book manuscript, forthcoming

[36] Bradley I and Price C (1991), 'Regulation through an average revenue constraint', *Regional Science and Urban Economics,* Vol 21, pp89-108

[37] Bradley I and Price C (1988), 'The economic regulation of private industries by price constraints', *The Journal of Industrial Economics,* Vol 37, pp99-106

[38] Braeutigam R and Panzar J (1993), 'Effects of the change from rate of return to price cap regulation', *American Economic Review,* Vol 83, pp191-198

[39] Braeutigam R and Panzar J (1989), 'Diversification incentives under 'price-based' and 'cost-based' regulation', *RAND Journal of Economics,* Vol 20, pp373-391

[40] Buckland R and Fraser P (2001), 'Political and Regulatory Risk: Beta Sensitivity in UK Electricity Distribution', *Journal of Regulatory Economics,* Vol 19, pp5-25

[41] Burns P, Turvey R and Weyman-Jones T (1995a), 'General properties of sliding scale regulation', *Technical Paper No 3, Centre for the Study of Regulated Industries*

[42] Burns P, Turvey R and Weyman-Jones T (1995b), 'Sliding scale regulation of monopoly enterprises', *Technical Paper No 11, Centre for the Study of Regulated Industries*

[43] Burns P and Weyman-Jones T (1996), 'Cost functions and cost efficiency in electricity distribution: a stochastic frontier approach', *Bulletin of Economic Research,* Vol 48, pp41-64

[44] Caillaud B, Guesnerie R, Rey P and Tirole J (1988), 'Government intervention in production and incentives theory: a review of recent contributions', *RAND Journal of Economics,* Vol 19, pp1-26

[45] Centre for Regulated Industries (various years), *The UK Water Industry - Charges for Water Services,* School of Management, University of Bath

[46] Centre for Regulated Industries (various years), *UK Electricity Industry Financial and Operating Review,* School of Management, University of Bath

[47] Coelli T (1996), 'A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program', CEPA Working Paper 96/08

[48] Coelli T, Rao D and Battese G (1998), *An Introduction to Efficiency and Productivity Analysis,* Kluwer Academic Publishers

[49] Competition Commission (2002), *Vodafone, O2, Orange and T-Mobilc: Reports on references under section 13 of the Telecommunications Act 1984 on the charges made by Vodafone, O2, Orange and T-Mobile for terminating calls from fixed and mobile networks,* HMSO

[50] Competition Commission (2000), *Mid Kent Water plc: A report on the references under sections 12 and 14 of the Water Industry Act 1991,* HMSO.

[51] Cowan S (1997), 'Tight average revenue regulation can be worse than no regulation', *The Journal of Industrial Economics,* Vol XLV, pp75-88

[52] Cox J and Isaac R (1987), 'Mechanisms for incentive regulation: theory and experiment', *RAND Journal of Economics,* Vol 18, pp348-359

[53] Crampes C and Estache A (1997), 'Regulatory trade-offs in the design of concession contracts', *World Bank Working Paper,* preliminary version

[54] Crew M (ed) (1994), *Incentive regulation for public utilities.* Topics in Regulatory Economics and Policy Series.

[55] Crew M and Kleindorfer P (2002), 'Regulatory Economics: Twenty Years of Progress', *Journal of Regulatory Economics,* Vol 21, pp5-22

[56] Crew M and Kleindorfer P (1996), 'Incentive regulation in the United Kingdom and the United States: Some Lessons', *Journal of Regulatory Economics,* Vol 9, pp211-225

[57] Currie D, Levine P and Rickman N (1999), 'Regulation and the ratchet effect: should regulators be pro-industry', *Regulation Initiative Discussion Paper Series No 26,* London Business School

[58] Dalen DM (2000), 'Catching-up investment without regulatory commitment', *Journal of Regulatory Economics,* Vol 18, pp133-150

[59] Dalen DM (1997), 'Regulation of quality and the ratchet effect; does unverifiability hurt the regulator?', *Journal of Regulatory Economics,* Vol 11, pp139-155

[60] De Fraja G and Iossa E (1998), 'Price caps and output floors: A comparison of simple regulatory rules', *The Economic Journal,* Vol 108, pp1404-1421

[61] Demski J and Sappington D (1984), 'Optimal incentive contracts with multiple agents', *Journal of Economic Theory,* Vol 33, pp152-171

[62] Demsetz H (1968), 'Why regulate utilities?', *Journal of Law and Economics,* Vol 11, pp55-65

[63] Department of the Environment (various years), *The Digest of Environmental Statistics,* HMSO

[64] Di Tella R and Dyck A (2002), 'Cost reductions, cost padding and stock market prices: the Chilean Experience with price cap regulation', *Harvard Business School Working Paper*

[65] Domah P and Pollitt MG (2001), 'The restructuring and privatisation of Electricity Distribution and Supply Businesses in England and Wales: A Social Cost-Benefit Analysis', *Fiscal Studies*, Vol 22, pp107-146

[66] dotecon (2001), *Estimation of fixed to mobile price elasticities: Prepared for BT.* Report available on www.dotecon.com.

[67] Drinking Water Inspectorate (various years), *Drinking Water Report by the Chief Inspector,* HMSO

[68] Electricity Association (2000), *Electricity Industry Review*

[69] Europe Economics and Professor Nick Crafts (1998), *Water and Sewerage Industries General Efficiency and Potential for Improvement,* Final Report prepared for Ofwat

[70] Fare R, Grosskopf S, Norris M and Zhang Z (1994), 'Productivity growth, technical progress, and efficiency changes in industrialised countries', *American Economic Review,* Vol 84, pp66-83

[71] Farrell MJ (1957), 'The measurement of technical efficiency', *Journal of the Royal Statistical Society. Series A (General).* Vol 120, pp253-290

[72] Finsinger J and Vogelsang I (1985), 'Strategic management behaviour under reward structures in a planned economy', *Quarterly Journal of Economics,* Vol 100, pp263-270

[73] Forsyth P (1999), 'Monopoly price regulation in Australia: assessing regulation so far', *Paper at the 1999 Industry Economics Conference*

[74] Freixas X, Guesnerie R and Tirole J (1985), 'Planning under incomplete information and the ratchet effect', *The Review of Economic Studies,* Vol 52, pp173-191

[75] Frontier Economics (2003a), *Developing Network Monopoly Price Controls: Workstream A. Regulatory Mechanisms for dealing with uncertainty. A final report prepared for Ofgem,* March.

[76] Frontier Economics (2003b), *Developing Network Monopoly Price Controls: Workstream B. Balancing Incentives. A final report prepared for Ofgem,* March.

[77] Gasmi F, Ivaldi M and Laffont JJ (1994), 'Rent extraction and incentives for efficiency in recent regulatory proposals', *Journal of Regulatory Economics,* Vol 6, pp151-175

[78] Gilbert R and Newbery DM (1994), 'The dynamic efficiency of regulatory constitutions', *RAND Journal of Economics,* Vol 25, pp538-554

[79] Glynn D (1992), *The Mechanisms of Price Control,* NERA Topics 7

[80] Gort M and Wall RA (1988), 'Foresight and public utility regulation', *The Journal of Political Economy,* Vol 96, pp177-188

[81] Granderson G and Linvill C (1996), 'The impact of regulation on productivity growth: an application to the transmission sector of the interstate natural gas industry', *Journal of Regulatory Economics,* Vol 10, pp291-306

[82] Green R and Haskel J (2001), 'Seeking a Premier League Economy: The Role of Privatisation', *Prepared for CEP/IFS/NBER Conference Seeking a Premier League Economy, London, December 7-8 2000.* Revised September 2001.

[83] Green JR and Stokey NL (1981), 'The value of information in the delegation problem', *Harvard Institute of Economic Research, Discussion Paper No 776*

[84] Halvorsen B and Larsen BM (1999), 'Changes in the pattern of households electricity demand over time', *Discussion Papers No 255, Statistics Norway Research Department*

[85] Harstad R and Crew M (1999), 'Franchise biding without holdups: utility regulation with efficient pricing and choice of provider', *Journal of Regulatory Economics,* Vol 15, pp141-163

[86] Hattori T, Jamasb T and Pollitt M (2002), 'A Comparison of UK and Japanese Electricity Distribution Performance 1985-1998: Lessons for Incentive Regulation', *DAE Working Paper WP0212,* Department of Applied Economics, University of Cambridge

[87] Helm D (2001), 'Making Britain More Competitive: A Critique of Regulation and Competition Policy', *Scottish Journal of Political Economy,* Vol 48, pp471-487

[88] Helm D (1995a), 'Introduction', in Helm D (ed), *British Utility Regulation: Principles, Experience and Reform,* The OXERA Press

[89] Helm D (1995b), 'British Utility Regulation: Theory, Practice and Reform', in Helm D (ed), *British Utility Regulation: Principles, Experience and Reform,* The OXERA Press

[90] Helm D and Rajah N (1994), 'Water regulation: The periodic review', *Fiscal Studies,* Vol 15, pp74-94

[91] Helm D and Yarrow G (1988), 'The Assessment: The Regulation of Utilities', *Oxford Review of Economic Policy,* Vol 4, ppi-xxxi

[92] Holmström B (1977), 'On Incentives and Control in Organisations, Chapter II: Delegation', *unpublished PhD dissertation,* Stanford GSB.

[93] Holmström B (1982), 'Moral hazard in teams', *The Bell Journal of Economics,* Vol 13, pp324-340

[94] Holmström B (1984), 'On the theory of delegation', Chapter 8 of Boyer M and Kihlstrom R (eds), *Bayesian Models in Economic Theory,* North-Holland

[95] Holmström B and Milgrom P (1991), 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design', *The Journal of Law, Economics and Organization,* Vol 7, pp24-52

[96] Holzleitner C (2001), 'Efficient cost passthrough', *Journal of Regulatory Economics,* Vol 20, pp91-97

[97] Houston G (1996), 'Regulatory Asset Valuation', *A briefing paper for the conference on Advanced Strategies for Company Valuation,* National Economic Research Associates

[98] Hsiao C (1986), *Analysis of Panel Data,* Econometric Society Monographs, Cambridge University Press

[99] Hulten CR (2000), *Total Factor Productivity: A Short Biography,* NBER Working Paper 7471, January

[100] Hunt LC and Lynk EL (1995), 'Privatisation and efficiency in the water industry: an empirical analysis', *Oxford Bulletin of Economics and Statistics*, Vol 57, pp371-388

[101] Jamasb T and Pollitt M (2000), 'Benchmarking and Regulation of Electricity Transmission and Distribution Utilities: Lessons from International Experience', Working Paper, Department of Applied Economics, University of Cambridge

[102] Jehiel P (1998), 'Repeated games and limited forecasting', *European Economics Review,* Vol 42, pp543-551

[103] Joskow P and Rose N (1989), 'The effects of economic regulation', in Schmalensee R and Willig R (eds), *Handbook of Industrial Organization, Volume II,* Elsevier Science Publishers, pp1449-1506

[104] Joskow P and Schmalensee R (1986), 'Incentive regulation for electric utilities', *Yale Journal on Regulation,* Vol 4, pp1-49

[105] Kahn AE (1988), *The Economics of Regulation: Principles and Institutions,* The MIT Press

[106] Kridel D, Sappington D and Weisman D (1996), 'The Effects of Incentive Regulation in the Telecommunications: A Survey', *Journal of Regulatory Economics,* Vol 9, pp269-306

[107] Kydland F and Prescott E (1977), 'Rules Rather than Discretion: The Inconsistency of Optimal Plans', *The Journal of Political Economy,* Vol 85, pp473-492

[108] Laffont JJ (1994), 'The new economics of regulation ten years after', *Econometrica,* Vol 62, pp507-537

[109] Laffont JJ and Martimort D (2002), *The Theory of Incentives: The Principal-Agent Model,* Princeton University Press

[110] Laffont JJ and Tirole J (2000), *Competition in Telecommunications,* MIT Press

[111] Laffont JJ and Tirole J (1993), *A Theory of Incentives in Procurement and Regulation,* MIT Press

[112] Laffont JJ and Tirole J (1990), 'Adverse selection and renegotiation in procurement', *The Review of Economic Studies,* Vol 57, pp597-625

[113] Laffont JJ and Tirole J (1988), 'The dynamics of incentive contracts', *Econometrica,* Vol 56, pp1153-1175

[114] Laffont JJ and Tirole J (1986), 'Using cost observations to regulate firms', *Journal of Political Economy,* Vol 94, pp614-641

[115] Lakshmanan TR and Anderson W (1980), 'Residential energy demand in the United States', *Regional Science and Urban Economics, Vol 10,* pp371-386.

[116] Legros P (1993), 'Information revelation in repeated delegation', *Games and Economic Behaviour,* Vol 5, pp98-117

[117] Levaggi R (1999), 'Optimal procurement contracts under a binding budget constraint', *Public Choice,* Vol 101, pp23-37

[118] Levine P (1999), 'Delegation and the Under-Investment Problem', *Regulation Initiative Discussion Paper No 32,* London Business School

[119] Lewis T and Sappington D (1997), 'Penalizing success in dynamic incentive contracts: no good deed goes unpunished?', *RAND Journal of Economics,* Vol 28, pp346-358

[120] Lewis T and Sappington D (1992), 'Incentives for Conservation and Quality-Improvement by Public Utilities', *The American Economic Review,* Vol 82, pp1321-1340

[121] Lewis T and Sappington D (1989), 'Regulatory options and price-cap regulation', *RAND Journal of Economics,* Vol 20, pp405-416

[122] Lewis T and Sappington D (1988a), 'Regulating and monopolist with unknown demand and cost functions', *RAND Journal of Economics,* Vol 19, pp438-457

[123] Lewis T and Sappington D (1988b), 'Regulating and monopolist with unknown demand', *The American Economic Review,* Vol 78, pp986-998

[124] Liesner (1995), 'The role of the MMC in utility regulation', in Helm D (ed), *British Utility Regulation: Principles, Experience and Reform,* The OXERA Press

[125] Littlechild S (2001), 'A review of UK electricity regulation 1999-2000', in Vass P (ed), *CRI Regulatory Review - Millennium Edition 2000/2001,* School of Management, University of Bath

[126] Littlechild S (1986), *Economic Regulation of Privatised Water Authorities: A Report to the Secretary of State,* Department of Environment: London.

[127] Littlechild S (1983), *Regulation of British Telecommunications' Profitability: A Report to the Secretary of State.* Department of Trade and Industry: London

[128] Liston C (1993), 'Price-cap versus rate-of-return regulation', *Journal of Regulatory Economics,* Vol 5, pp25-48

[129] Loeb M and Magat WA (1979), 'A decentralized method for utility regulation', *Journal of Law and Economics,* Vol 22, pp399-404

[130] Lowe P (1998), 'The reform of utility regulation in Britain: some current issues in historical perspective', *Journal of Economic Issues,* Vol 32, pp171-190

[131] Lyon T (1996), 'A model of sliding-scale regulation', *Journal of Regulatory Economics,* Vol 9, pp227-247

[132] Lynk EL (1993), 'Privatisation, Joint Production and the Comparative Efficiencies of Private and Public Ownership: The UK Water Industry Case', *Fiscal Studies,* Vol 14, pp98-116.

[133] Majumdar SK (1997), 'Incentive regulation and technical efficiency in the US Telecommunications Industry', *The Journal of Business,* Vol 70, pp547-576

[134] Mariñoso BG, Hviid M and Waddams Price C (2002), 'The Quality and Quantity of Regulatory Information', *CCR Working Paper CCR 02-5,* Centre for Competition & Regulation, UEA Norwich

[135] Markou E and Waddams Price C (1999), 'UK utilities: past reform and current proposals', *Annals of Public and Co-operative Economics,* Vol 70, pp371-416

[136] Mayer C (2000), 'Water: the 1999 price review', in Robinson C (ed), *Regulating Utilities: New Issues, New Solutions,* The Institute of Economic Affairs in association with the London Business School, Edward Elgar. Lectures from Autumn 1999.

[137] Mayer C and Vickers J (1996), 'Profit-Sharing Regulation: An Economic Appraisal', *Fiscal Studies,* Vol 17, pp83-101

[138] Monopolies and Mergers Commission (1997), *Northern Ireland Electricity plc: a report on a reference under article 15 of the Electricity (Northern Ireland) Order 1992,* HMSO

[139] Monopolies and Mergers Commission (1995a), *Scottish Hydro-Electric plc: a report on a reference under section 12 of the Electricity Act 1989,* HMSO

[140] Monopolies and Mergers Commission (1995b), *South West Water Services Ltd: a report on the determination of adjustment factors and infrastructure charges for SWW,* HMSO

[141] National Audit Office (2002), *Pipes and Wires: Report by the Comptroller and Auditor General,* HMSO

[142] National Grid Company (various years), *Annual regulatory accounts*

[143] National Grid Company (2001), *Seven Year Statement*

[144] Newbery D (1999), *Privatization, Restructuring and Regulation of Network Utilities,* The Walras-Pareto Lectures, at the École des Hautes Études Commerciales, Université de Lausanne, The MIT Press

[145] Newbery DM and Pollitt MG (1997), 'The restructuring and privatisation of Britain's CEGB - Was it worth it?', *The Journal of Industrial Economics,* Vol XLV, pp269-303

[146] Nickell S(1991), 'Biases in dynamic models with fixed effects', *Econometrica,* Vol 49, pp1399-1416

[147] OECD (1999), *The price of water: trends in OECD Countries.*

[148] Offer (1992a), *Electricity Price Controls Statement,* Press Release, February

[149] Offer (1992b), 'Future control on National Grid Company Prices', A Statement by the Director General of Electricity Supply, Press Release, July

[150] Offer (1993), *Electricity Distribution: Price control, Reliability and Customer Service - A Consultation Paper,* October

[151] Offer (1994), *The Distribution Price Control: Proposals,* August

[152] Offer (1995a), *REC Price Controls,* Press Release, March

[153] Offer (1995b), *The Distribution Price Control: Revised Proposals,* July

[154] Offer (1995c), *The Transmission Price Control Review of the National Grid Company,* Consultation Paper, November

[155] Offer (1996a), *The Transmission Price Control Review of the National Grid Company,* Second Consultation Paper, March

[156] Offer (1996b), *The Transmission Price Control Review of the National Grid Company,* Third Consultation Paper, May

[157] Offer (1996c), *The Transmission Price Control Review of the National Grid Company,* Fourth Consultation Paper, August

[158] Offer (1996d), *The Transmission Price Control Review of the National Grid Company: Proposals,* October

[159] Offer (1998a), *Review of Public Electricity Suppliers 1998 to 2000: Consultation Paper,* February

[160] Offer (1998b), *Review of Public Electricity Suppliers 1998 to 2000: Price Controls and Competition - Consultation Paper,* July

[161] Offer (1998c), *Review of Public Electricity Suppliers 1998 to 2000: PES Business Plans - Consultation Paper,* December

[162] Office of National Statistics (various years), *UK National Accounts,* HMSO

[163] Office of National Statistics (various years), *Economic Trends,* HMSO

[164] Ofgem (1999a), *Review of Public Electricity Suppliers 1998 to 2000: Distribution Price Control - Consultation Paper,* May

[165] Ofgem (1999b), *Distribution Price Control - Draft Proposals,* August

[166] Ofgem (1999c), *The distribution price control review and separation of businesses: Letter to the Chief Executives of PES Distribution Businesses,* October

[167] Ofgem (1999d), *Review of Public Electricity Suppliers 1998 to 2000: Distribution Price Control - Final Proposals,* December

[168] Ofgem (1999e), *Information and Incentives Project: Letter to the Chief Executives of the PES Distribution Businesses,* Press Release, December

[169] Ofgem (1999f), *The Transmission Price Control Review of the National Grid Company from 2001. Initial Consultation Document,* December

[170] Ofgem (2000a), *Information and Incentives Project: defining output measures and incentive regimes for PES distribution businesses: Update,* March

[171] Ofgem (2000b), *The Transmission Price Control Review of the National Grid Company from 2001. Initial Thoughts,* March

[172] Ofgem (2000c), *Information and Incentives Project: Output measures and monitoring delivery between reviews. Initial Proposals,* June

[173] Ofgem (2000d), *Information and Incentives Project: Public Workshop. Summary of Plenary Sessions,* June

[174] Ofgem (2000e), *The Transmission Price Control Review of the National Grid Company from 2001. Draft Proposals,* June

[175] Ofgem (2000f), *Information and Incentives Project: Review of PES Measurement Systems,* Report by PBPower for Ofgem, July

[176] Ofgem (2000g), *Review of NGC's operating cost efficiency for the 2002 to 2006 price control. A report by Arthur Anderson.* July.

[177] Ofgem (2000h), *Information and Incentives Project: Output measures and monitoring deliver between reviews: Final Proposals,* September

[178] Ofgem (2000i), *The Transmission Price Control Review of the National Grid Company from 2001: Transmission Asset Owner. Final Proposals,* September

[179] Ofgem (2000j), *Information and Incentives Project: Draft Regulatory Instructions and Guidance,* December

[180] Ofgem (2001a), *Information and Incentives Project: Incentive Schemes. Initial Thoughts,* January

[181] Ofgem (2001b), *Information and Incentives Project: Final Version of Regulatory Instructions and Guidance,* Letter to the PESs, February

[182] Ofgem (2001c), *Information and Incentives Project: Incentive Schemes. Update,* May

[183] Ofgem (2001d), *Information and Incentives Project: Summary of May 2001 IIP Workshop,* July

[184] Ofgem (2001e), *Information and Incentives Project: Incentive Schemes. Initial Proposals,* July

[185] Ofgem (2001f), *Information and Incentives Project: Developing the Incentive Scheme. Update,* November

[186] Ofgem (2001g), *Information and Incentives Project: Incentive Schemes. Final Proposals,* December

[187] Ofgem (2002), *Developing network monopoly price controls: Initial Consultation,* August

[188] Ofgem (2003a), *Developing network monopoly price controls: Update document,* February

[189] Ofgem (2003b), *Developing network monopoly price controls: Initial Conclusions,* June

[190] Ofgem (various years), *Report on distribution and transmission performance*

[191] Ofwat (1990), *Paying for Water: A time for decisions. A consultation paper issued by the Director General of Water Services on Future Charging Policy for Water and Sewerage Services,* November

[192] Ofwat (1991), *Cost of Capital: A Consultation Paper, Volumes 1 and 2*, July

[193] Ofwat (1992), *Assessing Capital Values at the periodic review: a consultation paper on the framework for reflecting reasonable returns on capital in price limits,* November

[194] Ofwat (1993a), *Paying for Growth. A consultation paper on the framework for reflecting the costs of providing for growth in charges*, February

[195] Ofwat (1993b), *Paying for Quality: The Political Perspective*, July

[196] Ofwat (1993c), *Setting Price Limits for Water and Sewerage Services: The framework and approach to the 1994 Periodic Review,* November

[197] Ofwat (1994), *Future Charges for Water and Sewerage Services: The outcome of the Periodic Review,* July

[198] Ofwat (1996), *1994 review of water company charging limits: the Periodic Review, Information Note No 17*, February

[199] Ofwat (1997a), *MD124: 1999 Periodic Review,* February

[200] Ofwat (1997b), *The proposed framework and approach to 1999 Periodic Review: A consultation paper,* June

[201] Ofwat (1998a), *Setting price limits for water and sewerage services: The framework and business planning process for the 1999 Periodic Review,* February

[202] Ofwat (1998b), *A proposed approach to assessing overall service to customers: A technical paper,* March

[203] Ofwat (1998c), *Open letter from the Director General of Water Services to the Secretary of State for the Environment, Transport and the Regions and the Secretary of State for Wales and accompanying paper: Setting the quality framework - an analysis of the main quality costings submission 2000-05,* April

[204] Ofwat (1998d), *Assessing the Scope for Future Improvement in Water Company Efficiency: A technical Paper,* April

[205] Ofwat (1998e), *Financial Model Rule Book: A technical paper,* October

[206] Ofwat (1998f), *Prospects for Prices: A consultation paper on strategic issues affecting future water bills,* October

[207] Ofwat (1999a), *Informing the final decisions on Raising the quality 2000-2005: An open letter to the Secretary of State for the Environment, Transport and the Regions and the Secretary of State for Wales,* January

[208] Ofwat (1999b), *Periodic Review Public Consultation: Executive Summary. Prepared by Research by Design,* March

[209] Ofwat (1999c), *MD145: The Framework for Setting Prices,* March

[210] Ofwat (1999d), *Draft Determinations: Future Water and Sewerage Charges 2000-05,* July

[211] Ofwat (1999e), *Final Determinations: Future Water and Sewerage Charges 2000-05,* November

[212] Ofwat (2003), *Memorandum of Understanding Between Ofwat and Watervoice,* March

[213] Ofwat (various years), *July Return Submissions*

[214] Ofwat (various years), *Report on company performance*

[215] Ofwat (various years), *Report on financial performance and capital investment of the water companies in England and Wales*

[216] Ofwat (various years), *Report on levels of service for the water industry in England and Wales*

[217] OXERA (2000), *Guide to the Economic Regulation of the Gas Industry,* The OXERA Press

[218] OXERA (1999), *Guide to the Economic Regulation of the Electricity Industry,* The OXERA Press

[219] OXERA (1998a), *Guide to the Economic Regulation of the Water Industry,* The OXERA Press

[220] OXERA (1998b), *Guide to Price Reviews at the MMC,* The OXERA Press

[221] Parker M (1996), 'General conclusions and lessons', in J Surrey (ed), *The British Electricity Experiment-Privatization: the Record, the Issues, the Lessons.* Earthscan Publications Ltd

[222] Pint E (1992), 'Price-cap versus rate-of-return regulation in a stochastic-cost model', *RAND Journal of Economics,* Vol 23, pp564-578

[223] Pollitt M (1999), 'A survey of the liberalisation of public enterprises in the UK since 1979', University of Cambridge Working Paper

[224] Rand Journal of Economics (1989), *Symposium on Price-Cap Regulation,* Vol 20

[225] Rees R and Vickers J (1995), 'RPI-X price-cap regulation', in Bishop M, Kay J and Mayer C (eds), *The Regulatory Challenge,* Oxford University Press

[226] Regional Electricity Companies (various years), *Annual regulatory accounts,* published separately by each company

[227] Resende M (1999), 'Productivity growth and regulation in US local telephony', *Information Economics and Policy,* Vol 11, pp23-44

[228] Riordan M (1984), 'On delegating price authority to a regulated firm', *Rand Journal of Economics,* Vol 15, pp108-115

[229] Robinson C (2000), 'Moving to a competitive market in water', *The Beesley Lectures on Regulation,* Series X

[230] Rovizzi L and Thompson D (1995), 'The regulation of product quality in the public utilities', in Bishop M, Kay J and Mayer C (eds), *The Regulatory Challenge,* Oxford University Press

[231] Roycroft T (1999), 'Alternative regulation and the efficiency of local exchange carriers: evidence from the Ameritech states', *Telecommunications Policy,* Vol 23, pp469-480

[232] Saal DS and Parker D (2001), 'Productivity and price performance in the privatised water and sewerage companies of England and Wales', *Journal of Regulatory Economics,* Vol 20, pp61-90

[233] Saal DS and Parker D (2000), 'The Impact of Privatisation and Regulation on the Water and Sewerage Industry in England and Wales: A Translog Cost Function Model', *Managerial and Decision Economics,* Vol 21, pp253-268

[234] Salant D and Woroch G (1992), 'Trigger price regulation', *RAND Journal of Economics,* Vol 23, pp29-51

[235] Sappington D (2000), 'Price regulation and incentives', in Cave M, Majumdar S and Vogelsang I (eds), *Handbook of Telecommunications Economics,* forthcoming.

[236] Sappington D and Sibley D (1988), 'Regulating without cost information: the incremental surplus subsidy scheme', *International Economic Review,* Vol 29, pp297-306

[237] Sawkins JW (2001), 'The development of competition in the English and Welsh Water and Sewerage Industry', *Fiscal Studies,* Vol 22, pp189-215

[238] Schmalensee R (1989), 'Good regulatory regimes', *RAND Journal of Economics,* Vol 20, pp417-436

[239] Shaoul J (1997), 'A Critical Financial Analysis of The Performance of Privatised Industries: The Case of the Water Industry In England and Wales', *Critical Perspectives on Accounting,* Vol 8, pp479-505

[240] Shleifer A (1985), 'A theory of yardstick competition', *Rand Journal of Economics,* Vol 16, pp319-327

[241] Sibley D (1989), 'Asymmetric information, incentives and price-cap regulation', *RAND Journal of Economics,* Vol 20, pp392-404

[242] Sobel J (1999), 'A reexamination of yardstick regulation', *Journal of Economics and Management Strategy,* Vol 8, p33-60

[243] Spence AM (1975), 'Monopoly, Quality and Regulation', *the Bell Journal of Economics,* Vol 6, pp417-429

[244] Staiger D and Stock JH (1997), 'Instrumental variables regression with weak instruments', *Econometrica,* Vol 65, pp557-586

[245] Surrey J (1996a), 'Introduction' in J Surrey (ed), *The British Electricity Experiment-Privatization: the Record, the Issues, the Lessons.* Earthscan Publications Ltd

[246] Szalay D (2001), 'Optimal delegation', Working Paper, *University of Mannheim*

[247] Thanassoulis E (2000), 'The use of data envelopment analysis in the regulation of UK water utilities: Water distribution', *European Journal of Operational Research*, Vol 126, pp436-453

[248] Tilley B and Weyman-Jones T (1999), 'Productivity growth and efficiency change in electricity distribution', *Presentation to the British Institute of Energy Economics Conference*, St John's College, Oxford, September

[249] Turvey R (2002), 'Price control of Distribution Networks', *Regulation Initiative Working Paper Series, No 51,* London Business School

[250] Uri N (2002), 'Assessing the Effect of Incentive Regulation on technical efficiency in Telecommunications in the United States', *European Journal of Law and Economics,* Vol 13, pp113-127

[251] Uri N (2001), 'Productivity Change, Technical Progress, and Efficiency Improvement in Telecommunications', *Review of Industrial Organization*, Vol 18, pp283-300

[252] Vickers J and Yarrow G (1988a), *Privatization: An Economic Analysis.* MIT Press: London, England and Cambridge Massachusetts

[253] Vickers J and Yarrow G (1988b), 'Regulation of privatised firms in Britain', *European Economic Review,* Vol 32, pp465-472

[254] Viehoff I (1995), *Evaluating RPI-X*, NERA Topics 17

[255] Vogelsang I (2002), 'Incentive Regulation and Competition in Public Utility Markets: A 20 Year Perspective', *Journal of Regulatory Economics,* Vol 22, pp5-27

[256] Waddams Price C (2002), 'Yardstick Competition and Comparative Performance Measures in Practice', in Hunt L (ed), Book in honour of Colin Robinson, Routledge, (forthcoming 2003)

[257] Waddams Price C (2000a), 'Efficiency and Productivity Studies in Incentive Regulation of UK Utilities', *Revista de Economia del Rosario*, Vol 3, pp11-24

[258] Waddams Price C (2000b), 'Gas: regulatory response to social needs', in Robinson C (ed), *Regulating Utilities: New Issues, New Solutions*, The Institute of Economic Affairs in association with the London Business School, Edward Elgar. Lectures from Autumn 1999.

[259] Waddams Price C (1998), 'The UK Gas Industry', in D Helm and T Jenkinson (eds), *Competition in Regulated Industries.* Oxford University Press: Oxford.

[260] Waddams Price C (1997a), 'Competition and Regulation in the UK Gas Industry', *Oxford Review of Economic Policy,* Vol 13, pp47-63

[261] Waddams Price C (1997b), 'Regulating for fairness', *New Economy,* Vol 4, pp117-122

[262] Waddams Price C, Brigham B and Fitzgerald L (2002), 'Service Quality in Regulated Monopolies', *CCR Working Paper CCR 02-4*, Centre for Competition & Regulation, UEA Norwich

[263] Waddams Price C and Hancock R (1997), 'UK privatization: effects on households', in Palma G and Sawyer M (eds), *Frontiers of Political Economy, Vol 7,* Routledge, pp68-79

[264] Waddams Price C and Weyman-Jones T (1996), 'Malmquist indices of productivity change in the UK gas industry before and after privatization', *Applied Economics,* Vol 28, pp29-39

[265] Water and sewerage companies (various years), *Annual regulatory accounts*, published separately by each company

[266] Waterson M (1995), 'Developing utility regulation in the UK', in Helm D (ed), *British Utility Regulation: Principles, Experience and Reform*, The OXERA Press

[267] Waterson M (1992), 'A comparative analysis of methods for regulating public utilities', *Metroeconmica*, Vol 43, pp205-225

[268] Weisman D (1994), 'Why less may be more under price-cap regulation', *Journal of Regulatory Economics,* Vol 6, pp339-362

[269] Weisman D (1993), 'Superior regulatory regimes in theory and practice', *Journal of Regulatory Economics,* Vol 5, pp355-366

[270] Weitzman ML (1980), 'The "Ratchet Principle" and Performance Incentives', *The Bell Journal of Economics,* Vol 11, pp302-308

[271] Weyman-Jones T (2001a), 'Stochastic Non-parametric Efficiency measurement and Yardstick Competition in Electricity Regulation', *Loughborough University, Economic Research Paper 01/10*

[272] Weyman-Jones T (2001b), 'Yardstick and Incentive Issues in UK Electricity Distribution Price Controls', *Fiscal Studies,* Vol 22, pp233-247

[273] Weyman-Jones T (1995), 'Problems of yardstick regulation in electricity distribution', in Bishop M, Kay J and Mayer C (eds), *The Regulatory Challenge,* Oxford University Press

[274] Williamson B (2001), 'UK 'Incentive Regulation': International Best Practice?' in P Vass (ed), *Regulatory Review 2000/2001 Millennium edition.* Centre for Regulated Industries, University of Bath.

[275] Williamson B (1997), *Incentives and Commitment in Regulation,* NERA Topics 20

[276] Wooldridge JM (2002), *Econometric Analysis of Cross Section and Panel Data,* The MIT Press

[277] WS Atkins Management Consultants in association with OXERA (2001), *External Efficiency Review of Utility Regulators,* Final Report for HM Treasury, February

[278] Yarrow G (1989), 'Regulatory Issues in the Electricity Supply Industry', in Helm D, Kay J and Thompson D (eds), *The market for energy,* Oxford University Press

[279] Yarrow G and Jasiñski P (eds) (1996), *Privatization: Critical Perspectives on the World Economy, Vols I-IV,* Routledge, London and New York

Appendix A

Background to the RPI-X game

In Chapter 2 we present a description of how RPI-X regulation works in the UK utility sectors. The background information used to create this description is summarised here[1], with the section headings corresponding to the relevant sections in Chapter 2. The information relates to our description of the players in the RPI-X game and to the description of decision-making in the Contract Agreement game. Our discussion of the regulated firm's decisions in the Implementation Game is based on standard economic analysis of a profit-maximising firm. This is because no public information is available about how the companies make their operating and regulatory decisions. We do, however, provide evidence on the extent of mid-period intervention by the regulators in the Implementation Game.

Our description is based on a review of the practice of RPI-X regulation in the following sectors: (1) electricity distribution; (2) electricity transmission; and (3) the water and sewerage sector. RPI-X regulation is also used in the gas transportation, airports, rail network and telecoms sectors. Secondary-source information about these sectors is occasionally used to validate statements made in the description. Inevitably with such a generalisation some of the details relate more to one sector than to others. In addition some details which are of importance in only one sector may not be mentioned here.

The information is sourced from consultation and decision documents published by the water regulator (Ofwat) and the energy regulator (Ofgem, formerly Offer). In the first round of reviews after privatisation little information was released on how the regulators came to their decisions. Ofwat was better than most, but the details of the final proposals for each firm were not published. Most of the facts have therefore been

[1] Our objective is to summarise the relevant information, and thereby save interested parties from having to digest the large amount of detail contained in the many reports published by Ofgem (Offer), Ofwat and the Competition Commission. A full list of the relevant reports is included in the bibliography for the interested reader.

taken from the most recent rounds of periodic reviews - ie, the 1999 water review, the 1999 electricity distribution review, and the 2000 electricity transmission review[2]. Due to an increased focus on openness and transparency, the regulators presented far more detail on how price caps were set in these reviews than ever before. Secondary information sources have also been used[3].

A.1 Overview of the RPI-X game

A.1.1 Playing a game

Chapter 2, and our discussion of the welfare properties of the RPI-X mechanism in Chapter 3, are based on the assumption that the regulator and the firm play an infinitely repeated regulatory game. This idea has been supported by others, including Vickers and Yarrow (1988a). Regulators have also indicated that they believe that a game is being played. For example, Ofgem (1999a) argues that companies place 'undue emphasis on the periodic process' and this, along with other elements of the regulatory regime, has 'led to a form of regulatory game between the regulator and the regulated companies'.

In each of the sectors, the regulator and the firm have met at regular intervals to agree on the next period's regulatory contract and, in some cases, to agree on changes in the contract mid-period. We present, in Tables A.1 to A.3, a summary of the main events which have led to these interactions since privatisation. These timelines support the idea of ongoing, and repeated, interactions between the players. We note that there would have been more interaction at recent price reviews, as the regulators have published an increasing number of consultation papers.

[2]A complementary description, and critique, of RPI-X regulation in the gas sector can be found in Waddams Price (1997a and 1998). Helm and Rajah (1994) provide an analysis of regulation in the water sector at the first periodic review, and Turvey (2002) presents a description of regulation in the electricity distribution sector. Littlechild (2001) also provides a review of recent regulatory developments in the electricity sector.

[3]Information has been taken from Armstrong, Cowan and Vickers (1994, Chs 9-11) and Vickers and Yarrow (1988a, Part II). Both books provide a detailed, but early, description of regulation in the gas, electricity, water and telecoms sectors. More recent information, particularly on the institutional framework in each of the sectors, was obtained from the OXERA guides on regulation in the water, gas and electricity sectors.

Table A.1: Timeline of events in the electricity distribution sector

1990	First price control came into effect for the 1990-1995 period (April 1st)
	The twelve RECs were floated on the stock exchange (Dec)
1993	1995 price review announced and consultation paper published (Oct)
1994	Price control proposals announced for 1995-2000 period (Aug)
	Proposals accepted by all 12 RECs
	Scottish Hydro-Electric rejected its price control proposals (Oct)
	Scottish Hydro-Electric case referred to the MMC (Nov)
1995	Re-review of price control announced (March)
	MMC published its decision on Hydro-Electric's price controls (June)
	Revised price controls announced for 1996-2000 period (July)
	Proposal accepted by all twelve RECs
	ScottishPower began a judicial review case against Offer
1996	NIE rejected the price control proposals presented by Ofreg (July)
	NIE case referred to the MMC (Sept)
	The High Court rejected ScottishPower's judicial review case (July)
	ScottishPower case referred to the Court of Appeal
1997	MMC published their decision on NIE's price controls (March)
	NIE rejected the regulator's revised contract (Aug)
	NIE case referred for judicial review (Sept)
	Court of Appeal ruled in favour of ScottishPower (Feb)
1998	1999 price review announced (Feb)
	Consultation papers published (Feb, July, Dec)
	The High Court rejected NIE's judicial review case (June)
	NIE case referred to the Court of Appeal (July)
	Court of Appeal rule in favour of NIE (Oct)
	Business plans submitted (Nov)
1999	Consultation papers published for 1999 price review (May, Oct)
	Draft price control proposals published (Aug)
	Final price control proposals published for 2000-2005 period (Dec)
	Proposals accepted by all 12 RECs
2000	Information and Incentives Project begins
	Utilities Act passed by Parliament

Table A.2: Timeline of events in the electricity transmission sector

1990	First price control came into effect for 1990-1993 period (April 1st)
1992	Price control proposals announced for 1993-1997 period (July)
	Prosposals accepted by NGC
1995	1996 price review announced and initial consultation published (Nov)
	NGC floated on the stock exchange (Dec)
1996	Consultation papers for 1996 price review published (March, May, Aug)
	Price control proposals published for 1997-2001 period (Oct)
	Proposals accepted by NGC
1999	2000 price review announced and initial consultation paper published (Dec)
2000	Business plan submitted (January)
	Consultation papers for 2000 price review published (March, Aug)
	Draft price control proposals published (June)
	Final price control proposals published for 2001-2006 period (Sept)
	Proposals accepted by NGC

Table A.3: Timeline of events in the water sector

1989	Water companies floated on the stock exchange (Dec)
1990	First price control came into effect for 1990-2000 period (April 1st)
1991	1994 price review announced and consultation paper published (July)
1992	Consultation paper for 1994 price review published (Nov)
1993	Consultation papers for 1994 price review published (Feb, July, Nov)
	Interim determinations to revise price caps of 17 companies
1994	Business plans submitted (March)
	Final price controls published for the 1995-2005 period (July)
	Proposals accepted by nine WASCs; rejected by one WASC and one WOC
	Two cases referred to the MMC (Sept)
1995	DG encourages firms to share efficiency savings early (April)
	MMC's recommendations published (July)
1996	DG, again, encourages firms to share efficiency savings early (July)
	1999 price review announced (Oct)
1997	Consultation papers for 1999 review published (Feb, June, July)
1998	Companies agree to change regulatory period to 5 years (January)
	Consultation papers for 1999 review published (Feb, March, April, July, Sept, Oct)
1999	Consultation papers for 1999 review published (Jan, March)
	Business plans submitted (April)
	Draft price control published (July)
	Final price controls published for 2000-2005 period (Nov)
	Proposals accepted by all 10 WASCs; rejected by two WOCs
2000	Cases referred to the Competition Commission (Feb)
	Commission recommendations are published (Sept)

We assume that the game is repeated at five-year intervals. This reflects the length of the regulatory periods in most of the utility sectors, as shown in Table A.4. Our assumption that there is no known end-date to the RPI-X game is justified by the details of each firm's licence. These licences are issued for an indefinite period and, hence, the firm is assumed to have an infinite right to the monopoly franchise. The Privatisation Acts include provisions for the relevant Secretary of State to remove a licence from the firm. This requires a long notice period, however, and notification could not be given until a minimum period after privatisation. For example, the notification period in the electricity sector is 25 years and notice could not be given until ten years after privatisation (ie, March 31st 2000). There are strict conditions under which notification to terminate a licence can be given - eg, continuous failure to provide sufficient capacity to meet demand. The regulatory agencies were also established indefinitely. While the individual Director General changes, the existence of the regulatory body, and its powers and duties, does not. The game is therefore expected to be played indefinitely.

We have assumed that the regulator makes all contract offers to the firm, and the firm chooses whether or not to accept them. This reflects the practice in the UK where only the regulator proposes changes to a contract[4],[5]. Under the Privatisation Acts the contract revision can only be implemented with the acceptance of the firm, or following a decision by the Commission. This places a constraint on the regulator.

[4]This is in direct contrast to the practice of rate of return regulation in the US where the firm often requested a rate hearing when costs increased beyond the current allowed level. The firms presented their own case on what the contract should be in these hearings.

[5]In the water sector a firm can request that the regulator undertake an Interim Determination during the period. In addition, under the ten-year regulatory period the firm could request that a periodic review be held after five-years. In both cases, while the firm has the right to request a contract review, it does not present a contract offer to the regulator. Once the review is underway, the regulator is responsible for designing the revised contract, and the game proceeds as described in Chapter 2.

Table A.4: Regulatory periods

Sector	Regulatory Period
Water[1]	1990-2000
	1995-2005
	2000-2005
Distribution	1990-1995
	1995-2000
	2000-2005
Transmission	1990-1993
	1993-1997
	1997-2001
	2001-2006
BT	1984-1989
	1989-1993
	1993-1997
	1997-2001
British Gas	1987-1992
	1992-1997
	1997-2002

Note: [1]The water industry was privatised with a ten-year regulatory period. The regulator had the option of revising the contract after five-years and this option was taken in both 1995 and 2000. In January 1998, the water industry agreed to a licence modification which formally changed the period to five-years.
Source: Ofgem, Ofwat and Offer - various publications; Armstrong, Cowan and Vickers (1994)

A.1.2 The Players

The regulated firm

We describe the regulated firm as a natural monopoly. This research therefore only applies to the following companies: the integrated water and sewerage companies (WASCs) and the water only companies (WOCs); the electricity distribution business of the Regional Electricity Companies (RECs); the electricity transmission company (NGC); the transportation business of British Gas (Transco); the railways operator (Network Rail, formerly Railtrack); and, the local-loop network business of British Telecom[6]. All of these firms have been regulated by the RPI-X mechanism since they were privatised. It is assumed that regulation will be required in the long-term given

[6]In the 2000 review, Ofgem made a clear distinction between NGC's role as a transmission assset owner (TO) and its role as a system operator (SO) on the network. This latter role has become increasingly important with the introduction of the New Electricity Trading Arrangements (NETA). We are interested in the long-term regulation of the core monopoly business, the transmission asset owner operations.

the absence of feasible competition in these sectors.

We have assumed that the regulated firm produces a single good. This is a simplification of reality. All the utilities provide a range of products to consumers. For example, electricity distribution companies distinguish between electricity supplied at different times of the day. Our description therefore abstracts from reality, and further research is required to explore the impact of the RPI-X game on the decisions of a multi-product firm.

We have also assumed that a number of regulated firms operate in the sector. This is only true for the electricity distribution and water sectors, where regional monopoly companies were established at privatisation. In the case of electricity transmission, telecoms, gas transportation and railways, the privatised network firm is a national monopoly. There is thus no scope - unless within company comparisons are possible - to introduce the relative scheme discussed in Chapter 3 in these sectors[7].

As noted above, the assumption that the firm is a profit-maximiser is made because we do not have any concrete evidence on what alternative objectives motivate the regulated firms. This assumption allows us to focus on the familiar economic analysis of profit-maximisation and cost-minimisation.

The economic regulator

Each sector has its own regulator agency, headed by a Director General (DG). In energy, the regulatory agency is called Ofgem. It was established in 1999 and represents the merger of the former gas regulatory agency (Ofgas) and electricity regulatory agency (Offer). The water regulatory agency is called Ofwat. The agencies are independent non-ministerial government departments and their running costs are financed through licence fees from the regulated industry.

The regulator is responsible for ensuring that the firm meets its licence conditions, and is able to propose changes to the licence where this is considered necessary. In particular, the regulator sets the firm's price cap, monitors the firm's charging regime to ensure that it meets the constraints of the cap, and proposes revisions to the cap over time. We note that the regulator is not required to use the RPI-X mechanism and could, at any time, change the form of regulation entirely (including deciding that no regulation is needed). In practice, the regulators in the natural monopoly sectors have adopted this form of regulation at every review since privatisation, and they continuously express their support for the price cap mechanism.

The regulator also sets quality of service standards for the firm and monitors the firm's performance relative to these standards. Increased powers in this area were

[7]Intra-company comparisons may be feasible in the gas sector. For example, cost comparisons can be made across Transco's Local Distribution Zones (LDZs).

given to the regulators under the Competition and Service (Utilities) Act 1992. No clear guidelines are given on the form of regulation which should be used for quality of service. All regulators now collect and publish annual information on quality of service performance. Guaranteed Standards Schemes exist which require the firm to pay compensation directly to customers who experienced service problems. In addition, since the 1999 price reviews, water and electricity distribution companies have faced the prospect of a service performance adjustment to the next period's price cap.

We have assumed that the regulatory agency can be modelled as an individual decision-making unit. In practice this was true up to 2000, with the individual DG in each sector operating as the primary decision-maker. There has been a move towards collective decision-making in recent years. In the Utilities Bill 2000, the government changed the organisational structure of the energy regulatory agency, Ofgem, so that it is now headed by a board (The Gas and Electricity Markets Authority) rather than by an individual. The chairman of this board operates like the DG in terms of making public statements on behalf of the board but final decisions are made collectively rather than by a single individual. The newly created regulator in the communications sector, Ofcom, will also be headed by a board with both a chairman and a chief executive. Implications of this collective approach to decision-making should be examined in future research.

In our description of the RPI-X game the regulator makes decisions which reflect his current position on the trade-off between allocative efficiency, technical efficiency and distribution. This is our *interpretation* of how the regulator translates his legal powers and duties into a set of objectives.

The regulator's duties are set out in sectoral legislation and can be summarised as follows.

- Protect consumer interests (long-term and short-term).
- Ensure the firm carries out its functions properly.
- Set price limits which allow the firm to finance the proper carrying out of its functions[8].

In the water sector the duty to protect consumer interests is secondary to the financing of functions duty. In contrast, in the energy sector, the regulator's 'principal statutory objective is to protect the interests of consumers (present and future)'[9].

[8]In the water sector there is an additional requirement that the financing of functions duty must be carried out to enable the firm to earn an adequate rate of return on its investment. This introduces an element of rate of return regulation into the price cap regime.

[9]Ofgem (2002).

The regulator has a significant amount of discretion in interpreting these duties and deciding how they should be delivered. This regulatory discretion exists because the Privatisation Acts established very high-level duties but provided no guidance on how they should be interpreted or implemented.

We have also assumed that the regulator only considers the impact of his decisions on the current period (ie, the next five years). This assumption is justified for two reasons. First, regulators are appointed for five-year periods, usually to coincide with the regulatory period. There is the option for the post to be renewed (as occurred for Ian Byatt, the first water regulator, and Prof Stephen Littlechild, the first electricity regulator) but this cannot be presumed. The second reason why the regulator only considers the short-term effect of his decisions is that he is, legally, unable to commit a future regulator to any decisions or methodologies used. The implication of this is that their is no point in the regulator making decisions about the medium- to long-term as these will not necessarily be implemented. This is also the reason why we assume that only short-term contracts are feasible in the RPI-X game.

Competition Commission

The Competition Commission is an independent organisation which undertakes investigations into anti-competitive or monopolistic behaviour when cases are referred to it by the Director General of Fair Trading or by the Secretary of State for Trade and Industry. In addition, the Commission is required to act as an appeals body in the utility sectors when a firm rejects a proposed revision to its licence. It is this role which is of interest to us.

The Commission's stated objective is to ensure that the price cap which is included in the licence is in the public interest. There is no clear definition of what this means however. The Commission is required to take account of the regulator's duties when setting a price cap. These may vary slightly from one sector to the next[10]. This suggests that the objective function for choosing one price cap over another will be similar to that used by the regulator.

Each case is reviewed on its own merits and decisions made are expected to reflect the information supplied during the inquiry. This means that the Commission may have a different information set to the regulator. It also means that the Commission may make decisions which are only appropriate for the specific firm in question. These decisions do not apply to the rest of the industry, even if they are generally relevant (eg, a decision on the value of the cost of capital).

[10]Details of how the Commission interprets the regulator's duties can be found in the reports published for each case. The relevant references are provided in the bibliography.

The Commission's public interest decision and recommendations are published in a detailed final report. The regulator then devises an alternative contract, taking account of these recommendations, and this is implemented in the firm's licence. The Commission has no further role in the process. In the water sector, the regulator is obliged to implement the Commission's recommendations exactly. In the energy sector, however, the regulator has discretion in deciding on the final contract. The current price cap must be amended if the Commission found it to not be in the public interest. The licence change is expected to deal with the specific concerns raised by the Commission, but it does not have to exactly reflect the Commission's recommendations.

Non-economic regulators

A number of other regulators are involved with the regulation of the utility sectors.

- In the electricity sector, the Health & Safety Executive (HSE) is responsible for ensuring that the regulated firms meet legal standards relating to safety on the network.
- The Environment Agency (EA) is responsible for setting the environmental standards which water companies must meet when they discharge sewage to coastal waters, rivers and lakes. The EA is also responsible for the national management of water resources.
- The Drinking Water Inspectorate (DWI) is responsible for regulating the quality of water at the consumer's tap.

All these organisations contribute to the price-setting process by presenting information to the economic regulator on the outputs which are required for the next regulatory period. The water sector regulators are most active as the costs of environmental and drinking water improvements have a major impact on the price control. The non-economic regulators also work, on an on-going basis, to ensure that firms are meeting their legal duties with respect to output delivery. They are responsible for ensuring that the firm introduces appropriate remedies when problems arise.

The government

The government influences the operating environment in which the regulator and the firm make their decisions. The Secretary of State responsible for each of the sectors can:

- terminate the licences of existing firms, in exceptional circumstances, and appoint new firms in their place;
- remove the existing DG, in exceptional circumstances or at the end of his contract period, and appoint a new DG;
- provide guidance to the regulator on statutory obligations (usually environmental and social) which the firm must meet;
- set financial penalties for breach of statutory obligations; and
- introduce new legislation relating to the regulatory institutions and the market structure.

When making decisions the Secretary of State is bound by the same duties as the regulator.

The government has also, through taxation, directly affected the profits earned by the regulated firms. Most notably, in the 1997 Budget the Labour Government introduced a Windfall Tax on utility profits. In this way, the government by-passed the regulator to share the efficiency savings made by firms with the general public.

By setting the initial price caps at privatisation the government set a precedent of what was expected from price cap regulation. These initial caps influenced the decisions which regulators had to make in the future. For example, Surrey (1996a) emphasises that 'We repeatedly identify the importance of the very lenient initial settlement which was determined by the Government (as opposed to the regulator) as the prime cause of the imbalance between the benefits received so far by shareholders on the one hand and consumers on the other'. There is thus a sense in which future price reviews needed to 'correct' for the leniency of the caps set at privatisation.

Consumer groups

A number of consumer organisations operate in the utility sectors.

- In the energy sector customers are represented by energywatch. This organisation, established under the Utilities Act 2000, represents the merger of the Gas Consumer Council and the Electricity Consumers' Committees which were established at privatisation. energywatch is an independent body which works with, but entirely separate from, the energy regulator.
- The water consumer organisation, WaterVoice, was established in April 2002. It represents a name change for the Ofwat National Consumer Council, the

umbrella organisation of the Ofwat Customer Service Committees. These organisations were established at privatisation. WaterVoice is currently a part of Ofwat and its resources are financed by the regulator. It is due to be replaced by an independent Consumer Council for Water in the next Water Act[11]. We assume that the consumer group retains an arms-length relationship from the water regulator and, in particular, that it is entitled to hold its own views.

- The National Consumer Council represents consumers in all areas of the economy, including the utility sectors.

These consumer groups operate at both regional and national level. This ensures that firm-specific matters are dealt with, and that a consistent position is taken on industry-wide concerns. The groups investigate complaints made about individual companies, provide free information to consumers about the industry and their local service provider, and undertake independent monitoring of each firm's output delivery and charging performance. They also represent consumer interests in discussions with the economic and non-economic regulators, with the relevant industry, and with the government and the European Commission. A key role of the consumer groups is to provide the regulator with information on customer complaints, and on consumer preferences with respect to prices and output delivery.

A.2 The contract agreement game

A.2.1 The regulator's move

We assume that the RPI-X mechanism applies to the unit price of the single good that the firm produces. In practice, as noted above, the regulated firms produce a number of products and the cap is applied to an index of regulated prices[12]. In the water and airports sectors, the cap is applied to a tariff basket which is equal to the weighted sum of the prices of individual regulated products. The weights are calculated as the proportion of total revenue attributed to each product in the previous year. These are fixed from the current period forward but are endogenous over time. In the electricity distribution and gas sectors, the cap applies to changes in average revenue earned

[11] It is now expected that the Council will not be established before April 2005 (see Ofwat 2003).

[12] Armstrong, Cowan and Vickers (1994) note that the regulation of an index of prices is preferable to the regulation of prices separately as 'it gives the firm more freedom to respond to uncertainty in its costs'. There is no cap set on individual prices but prices are subject to standard restrictions implied by competition legislation (eg, prices must be non-discriminatory). In addition, rebalancing across prices might be restricted by universal service obligations and other limits on cross-subsidisation. A key element of the regulator's role, which is not discussed here, is deciding which prices should be included in the regulated index and what form of index to use.

from all regulated products. In electricity transmission, the cap applies to changes in maximum total revenue earned from regulated products. The cap on all these different price indices operates in the way described in Chapter 2 for a single price.

In our description, the regulator determines the cap by mechanically calculating the net present value of required revenues for the period. This broadly reflects the approach taken by the regulators since privatisation[13]. The final price cap determination does not generally fall out of a financial model, and is likely to be chosen on the basis of the regulator's judgement about the appropriate profile of prices for the period. Regulators do use financial models to calculate price cap *ranges*, however. They do not publish the actual financial models used but they emphasise the 'building block' approach by publishing detailed consultation papers which cover all of the elements of the allowed revenue calculation. The reports published by the Commission also show the type of modelling which is used to determine allowed revenues. In addition, Ofwat (1998e) published its Financial Rule Book for the 1999 periodic review. This provided more detailed information on the modelling approach used.

The firm is not constrained to make cost or financing decisions which lie within the assumptions made by the regulator. Ofgem (1999b) stressed that the capital projections used to set the price control 'do not represent a fixed sum to be spent and can only be likened to a contract in the sense that the price control is effectively a fixed price contract for the delivery of output'. Similarly, Ofwat (1999e) emphasised that 'Companies now know exactly what they should achieve, and the price limits which will enable them to deliver outcomes for customers and for the environment. Ofwat and the quality regulators will focus on the outcomes. Delivery by companies will be monitored on an annual basis. How they deliver is their responsibility; they have scope for innovative and cost-effective solutions'.

We assume that the regulator has asymmetric information about the firm's future costs, and that he observes historical total cost information from all firms in the sectors. Under the terms of its licence, the firm is required to present the regulator with all information requested, both historical and forecast. In addition, the firm is required under corporate legislation to publish annual financial accounts for the regulated business. The regulator therefore has access to a wide range of financial and non-financial data. Table A.5 summarises the type of information that is used by the regulators. Comparator information is only available to the regulators in the water and electricity distribution sectors. In other sectors, comparisons are often made with firms in other regulated sectors.

[13] Armstrong, Cowan and Vickers (1994) and Beesley (1996a) provide discussions of how this net present value approach is used in the regulated sectors.

Table A.5: The regulator's information

Demand	
Knows	Demand in next period
	Demand in previous periods
Forecasts	Long-term demand
Info for forecast	Firm's demand forecast
	Historical trends
Capacity	
Knows	Existing capacity
Forecasts	Required capacity
Info for forecast	Expected change in demand
Operating costs	
Knows	Level in previous periods - all firms in sector
Forecasts	Operating effort level in next period
Info for forecast	Business plan - all firms in sector
	Historical trends - all firms in sector
	Productivity in comparator industries
Capital investment	
Knows	Level in previous periods - all firms in sector
	Output targets in previous periods
Forecasts	Capital productivity
	Capital level in next period
Info for forecasts	Business plan - all firms in sector
	Historical trends - all firms in sector
	Productivity in comparator industries
Maintenance	
Knows	Expenditure level - all firms in sector
	Serviceability record - all firms in sector
Forecasts	Required expenditure
	Risk of future service problems
Info for forecasts	Firm's report on asset condition
	Historic maintenance levels - all firms in sector
	Serviceability record - all firms in sector
	Forecasts - all firms in sector
Cost of capital	
Knows	Interest rates, equity risk premium, betas and debt premium
Forecasts	Cost of capital for next period
Info for forecasts	Market information + academic advice
Outputs	
Knows	Output targets in previous periods - all firms in sector
	Output delivered in previous periods - all firms in sector
Forecasts	Impact of revised target on capital investment
Info for forecasts	Historic per unit cost levels

The problem faced by the regulator is that he does not have the technical expertise to determine how total cost information is affected by current technology, the firm's level of effort, the amount of output produced, and exogenous factors. In addition, the regulator is unable to determine whether the accounting information, or information on output delivery, is accurate. Auditors and the non-economic regulators are used as checks here, but concerns remain that the firm may alter the historical data so as to improve its position with the regulator. Similarly, as discussed in Chapter 2, the regulator cannot determine whether the forecasts provided by the firm are honest. The regulator therefore has access to a relatively large data set but problems exist with the quality of the data, and with the regulator's ability to interpret the information appropriately when making his own forecasts.

Our assumption that demand is known with certainty is limiting. In practice the regulator and the firm often use the same demand forecasts provided by independent organisations, or the regulator uses the firm's own forecasts. In this sense there is no asymmetry of information about demand, although there is uncertainty about how it evolves over time. Similarly, our assumption that the firm is certain about how its costs change over time is unrealistic. In practice the firm will have some uncertainty about how the operating environment, and hence its costs, will change during the regulatory period. For example, environmental standards may be tightened or there may be a reduction in the interest rate at which capital is financed. We do not consider this uncertainty in our description so that we can focus on the impact of asymmetric information between the regulator and the firm. The reader is referred to Frontier Economics (2003a) for a useful analysis of how the energy regulator might design regulatory mechanisms to deal with this uncertainty.

Sharing past cost savings

We assume that the regulator shares historical efficiency savings with consumers at the periodic review. Littlechild (1986) recognised that the X factor would need to be revised periodically in the sectors where there was little prospect of competition developing. The rules used have reflected the regulators' judgements about the appropriate trade-off between improved distribution and reduced technical efficiency. The absence of a formal theoretical model for sharing savings meant that the rules varied by sector, and they were changed by regulators within each sector over time[14]. This is why we assume, in our description of the RPI-X game, that the regulator uses his discretion when setting the rule, and why we assume that there is no commitment to the rule over time.

[14]Houston (1996) provides a useful summary of the sharing rules used by the regulators in the early years after privatisation.

A brief overview of sharing rules shows how the priority placed on different objectives has changed over time. At privatisation the government was concerned with ensuring that the firm was profitable (to ensure a successful privatisation), and that it improved its technical efficiency under the regulatory regime. The technical efficiency objective therefore dominated the distribution objective at this point, and there was no discussion about sharing the firm's 'efficiency rent' with consumers. After some time, however, the regulated firms earned high profits, and the regulators were placed under pressure by the government and consumer groups to share some of these profits with consumers. At this point the distribution objective became more important, and the regulators introduced large price cuts across the sectors. This reduced the net present value of the return earned by the firm and thereby dulled the incentive to improve technical efficiency. We see that the regulator's judgement about the appropriate sharing rule changes over time, reflecting his changing position on the importance of the distribution objective relative to the technical efficiency objectives.

Table A.6 provides a number of examples of how the regulators shared operating cost savings. We see that the length of the regulatory lag has been reduced to equal the length of the regulatory period in most sectors. In cases where the regulatory lag was longer than the regulatory period, the regulators used a glidepath approach to share savings gradually with consumers.

These examples show that the regulator makes no commitment, from one review to the next, as to what the operating cost sharing rule will be. This reflects the degree of discretion and judgement which is used to set the rule in the first place. For example, during the 1994 electricity distribution review, Offer (1994) stated that 'in future I would not expect the same need to make one-off changes in price, so there would be a correspondingly stronger case for incorporating all appropriate adjustments into the value of X'. This suggests that savings would be passed onto customers more gradually in the future. Contrary to this statement, Ofgem did deliver another large one-off price cut to customers in the 1999 periodic review. Similarly, in the 1994 price review Ofwat allowed water companies to retain efficiency savings made in 1990 to 1995 for 10 years (ie, to 2005). In the 1999 price review the regulatory lag was changed to 5 years, and savings made in 1995/96 were returned in the 2000/01 price cap (ie, five years earlier than originally proposed).

Table A.6: Operating Cost Sharing Rules

Review	Rule	Implied λ_o^{ij}
Water		
1995 review	Lag=10 years	$1 \, \forall \, i \leq j-2$
	Glidepath	$\frac{\widehat{e}^i}{e^i} < \lambda_o^{ij} < 1, \; i = j-1$
1999 review	Lag=5 years	$1 \, \forall \, i$
	Rolling mechanism	
BT		
1992 review	Lag=10 years	$1 \, \forall \, i \leq j-2$
	Glidepath	$\frac{\widehat{e}^i}{e^i} < \lambda_o^{ij} < 1, \; i = j-1$
Manchester Airport		
1997 review (MMC)	Lag=10 years	$1 \, \forall \, i \leq j-2$
	Glidepath	$\frac{\widehat{e}^i}{e^i} < \lambda_o^{ij} < 1, \; i = j-1$
NGC		
1997 review	Lag=5 years	$1 \, \forall \, i$
British Gas		
1997 review	Lag=5 years	$1 \, \forall \, i$
RECs		
1995 review	Lag=5 years	$1 \, \forall \, i$
1999 review	Lag=5 years	$1 \, \forall \, i$

Source: Competition Commission, Monopolies and Mergers Commission, Ofgem, Ofwat and Offer - various publications

Regulators have not varied their decisions on sharing capital savings very much. The trend has gone from retaining capital forecasts in the regulatory capital value indefinitely, to replacing forecast capital costs with actual capital costs at the end of each regulatory period. A number of examples are given in Table A.7. The level of actual investment included in the regulatory capital value may higher than the allowed level, but lower than the firm's actual capital costs. This is emphasised by Frontier Economics (2003a) who note that 'the regulator has discretion over whether or not to include items of historical capital expenditure' when adjusting the regulatory capital value at each review. Ofgem (1999d) include higher actual levels in the opening regulatory capital value if they are deemed to be 'necessary and efficient' or 'prudent'. In the water sector, the regulator undertakes a process of 'logging up' when investment increases during the period because of changes in environmental or drinking water outputs.

Table A.7: Sharing rules for capital investment

Review	Investment in RCV	Implied λ_c^{ij}
RECs		
1999 review	Actual 1991-1999	$\frac{\left(I^i+S^i+N^i-\delta^i\right)}{\left(\widehat{I^i}+\widehat{S^i}+\widehat{N^i}-\widehat{\delta^i}\right)}$
	Actual 2000-2005 (rolling)	$\frac{\left(I^i+S^i+N^i-\delta^i\right)}{\left(\widehat{I^i}+\widehat{S^i}+\widehat{N^i}-\widehat{\delta^i}\right)}$
NGC		
1996	Actual 1990-1996	$\frac{\left(I^i+S^i+N^i-\delta^i\right)}{\left(\widehat{I^i}+\widehat{S^i}+\widehat{N^i}-\widehat{\delta^i}\right)}$
2000	Actual 1997-2000	$\frac{\left(I^i+S^i+N^i-\delta^i\right)}{\left(\widehat{I^i}+\widehat{S^i}+\widehat{N^i}-\widehat{\delta^i}\right)}$
WASCs		
1994	Forecast 1990-1994	1
1999	Actual 1990-2000 (rolling)	$\frac{\left(I^i+S^i+N^i-\delta^i\right)}{\left(\widehat{I^i}+\widehat{S^i}+\widehat{N^i}-\widehat{\delta^i}\right)}$

Source: Ofgem, Ofwat and Offer - various publications

In our description of the RPI-X game we do not examine cost levels in each year of the regulatory period. In practice, the sharing of cost savings at the periodic review creates a timing distortion in the firm's incentive to make savings during the period. The firm has an incentive to make higher cost savings in the early years of the regulatory period. This is because any savings made at this point are retained for the full period, while savings made later are also only retained until the next periodic review, and hence for a shorter number of years. This effect was noted by Arthur Anderson (Ofgem 2000g) when examining trends in NGC's costs.

In the 1999 water review Ofwat introduced a rolling mechanism for passing on past operating and capital cost savings to customers. Under this methodology savings are retained by the firm for five years, independent of when the savings are made. These rolling rules are intended to smooth the firm's efficiency incentives over the regulatory period. Ofgem also made a commitment, in the 1999 distribution review, that at the next periodic review asset values would be adjusted on a five-year rolling basis. It is not clear how credible this commitment is as we know the regulator cannot require another regulator to use this methodology.

As an aside, we note that Ofwat (1997) explicitly recognised the impact that the sharing rules have on the firm's efficiency rent. Calculations were provided showing the impact of different operating cost sharing rules on the firm's efficiency share. Under the 1994 glidepath approach, where outperformance was transferred to customers progressively over a ten year period, the firm obtains 32% of the net present value of the savings made. Under a five-year straight line glidepath, where outperformance is transferred to customers gradually over a five-year period, the firm earns 28% of the net present value of savings. If all savings are shared at the start of the next regulatory period the firm gets 18% of the net present value of savings. These examples illustrate the impact of the regulator's sharing rule on the efficiency rent earned by the firm. Frontier Economics (2003b) and Williamson (2001) provide further examples of the impact of different rules on the firm's efficiency share.

Forecast operating costs

In our description, we assumed that the regulator sets the firm's base operating costs equal to its current operating cost level, adjusted downwards to reflect the regulator's assumption of what the base costs of an efficient firm would be given the current technology. Table A.8 provides examples of the methodologies used to set base operating costs by the water and electricity regulators in recent reviews. When making these adjustments the regulators take account of the impact of firm specific-factors, such as operating conditions and the state of the assets inherited at privatisation, on total costs. For example, Offer (1994) emphasised that its cost forecasts were based on the assumption that companies 'were managed efficiently and that they maintained or improved standards of service, but taking account of the particular circumstances under which each company had to operate'.

Table A.8: Efficient base operating cost adjustments

Review	Efficient benchmark	Firm's base operating costs
RECs		
1994	Average industry costs in 1992/93	Benchmark at start of period
1995		Removed redundancy costs
1999	Normalised average costs in 1997/98	Standardised own costs[1]
NGC		
2000	Efficient costs[2]	Benchmark at start of period
WASCs		
1994	Average industry costs in 1992/93	Actual adjusted for exceptional costs
1999	Average industry costs in 1997/98	Actual adjusted for exceptional costs

Note: [1]The firm's own operating costs were normalised to bring them in line with average costs in the industry. Adjustments were made for differences in accounting policies, cost allocation decisions and firm-specific factors. The regulator also reduced the firm's own costs to bring them in line with the efficient industry benchmark.

[2]The efficient cost benchmark was calculated by examining NGC's own historical costs, the costs of comparator utility companies (including the Public Electricity Supply Companies), the costs of transmission companies in other countries and the costs of non-utility companies.

Source: Ofgem, Ofwat and Offer - various publications

The regulator rolls the efficient base operating costs forward to reflect his assumption about the expected improvement in the firm's productivity level. We assume that the productivity targets incorporate a shift in the industry frontier and an additional firm-specific 'catch-up' element. Table A.9 provides examples of the information used by regulators to determine the size of the shift in the industry frontier.

Table A.9: Sectoral efficiency targets

Review	Basis for industry target
RECs	
1994	Evidence from other industries
	Information from companies
	Advice from consultants
	Best operating practice
1999	Average productivity growth in UK economy
	Advice from consultants
NGC	
1996	Business plan costs
	Information on 'well-managed' companies
	International comparisons
	Advice from consultants
2000	Business plan costs
	Costs of other utilities
	Advice from consultants
WASCs	
1994	Performance in UK economy
1999	Costs in international water sectors
	Costs of other UK sectors

Source: Ofwat, Offer and Ofgem - various publications

Detailed methodologies are used to calculate the firm-specific element of the productivity targets in the water and electricity distribution sectors. According to Armstrong, Cowan and Vickers (1994) the Department of the Environment undertook 'a comparative efficiency review' when setting the K-factors at the water sector privatisation. Ofwat used econometric analysis to compare the firms' efficiency levels in the 1994 and 1999 water price reviews. In the 1994 and 1999 distribution reviews, Ofgem (Offer) used regression analysis to determine whether each company was on or below the efficiency frontier[15]. The frontier was established by the base costs of a peer group. Ofgem combined the regression analysis with work provided by consultancy firms. It was stressed that the regulator was wary of placing 'an undue reliance on a statistical analysis of operating costs' (Ofgem, 1999d).

Both Ofgem and Ofwat use the comparator analysis to band firms into relative efficiency levels. They shy away from setting a different target for each firm because of concerns about the limitations of econometric analysis in precisely defining the relative efficiency of individual firms. The banding of firms is often influenced by other factors

[15] Weyman-Jones (2001a and 2001b) provides a critique of the methodology used by Ofgem in the 1999 price review.

such as advice from consultants, output delivery performance, and company-specific circumstances.

In the electricity distribution sector, the regulator set the same efficiency target for all firms, but used the efficiency banding to determine firm-specific adjustments to efficient base operating costs. In the water sector, the efficiency bands are used to set different efficiency targets for the firms over the entire period, and base operating costs are based on the firm's own costs.

Table A.10 shows the operating efficiency targets used by regulators in the water and electricity distribution sectors. The range of targets across the industry can be quite wide. We also see that the targets have tightened over time and inefficient firms are being required to catch-up with the frontier early in the period.

Table A.10: Firm-specific efficiency targets

Review	Catch-up decision	Target range
RECs		
1994	In base operating cost adjustment	2% pa
1999	75% of the way to frontier by 2001/02	2.3% pa
	Move in line with frontier to 2004/05[1]	
NGC		
1996	n.a.	2.5% pa
1999	n.a.	3.5% pa
WASCs		
1994	50% of the way to frontier by 2000	Water: 1-3.5% pa
		Sewerage: 1-3.4% pa
1999	Meet most efficient firm early in period	Water: 1.4-4.5% pa
		Sewerage: 1.4-5% pa

Note: [1]Ofgem assumed that the industry frontier for electricity distribution would move in line with the RPI for the control period. That is, the expected productivity improvement would be no greater than that for the economy in general.

Source: Ofwat, Offer and Ofgem - various publications

Regulatory capital value

We assume that the opening regulatory capital value is calculated using information on the firm's market value at privatisation. Table A.11 shows how this value is calculated in the water and electricity sectors. We see that it is essentially the actual market value of the company at flotation adjusted for sectoral-specific factors.

Table A.11: Opening regulatory capital value

Review	Methodology
RECs	
1994	Market value at flotation plus a 50% uplift
1995	Market value at flotation plus a 15% uplift
1999	As for 1995
NGC	
1996	Market value at flotation plus adjustment for Energis
2000	As for 1996
WASCs	
Privatisation	Discounted current cost cashflows from existing assets
1994	Average market value over first 200 days trading
1999	As for 1994

Source: Armstrong, Cowan and Vickers (1994); Ofgem, Offer and Ofwat - various publications

The opening value at flotation is rolled forward by actual net new investment to give the opening value at the start of the next regulatory period. The regulator then forecasts required investment for the next regulatory period to determine the forecast asset value.

We assume that regulators use the firm's capital cost forecasts as the basis for assessing required net new investment. Table A.12 provides examples of the size of adjustments made by individual regulators to these forecasts. The rationale for the adjustments vary by sector but in general they follow those outlined in our description of the game. In particular, consultants are used to undertake detailed benchmarking analysis to ensure that the forecast investment levels are efficient relative to the best available technologies. Where the firm's forecasts are considered too high, relative to the efficiency benchmark, downward adjustments are made.

Table A.12: Adjustments to capital investment forecasts

Review	Forecast	Adjustment
RECs		
1994	£5.6bn	-£660m
1999	£6bn	-£950m
NGC		
1996	£1.1bn	-£700m
2000	£1.5bn	-£220m
WASCs		
1999	n.a.	-£655m[1]

Note: [1]This is the value of the WASCs quality improvement programme which was disallowed by the regulator because the programmes proposed by individual companies were not supported by the DWI or EA, and/or because the regulator deemed the proposed solutions to not be 'cost-effective'.
Source: Ofgem, Ofwat and Offer - various publications

The regulators also make downward adjustments to the forecasts because their assessment of capital effort is higher than that used by the firm. The sectoral productivity target is determined by examining changes in capital productivity in other sectors in the economy, and in the same sector in other countries. The regulators also take account of any expected technology shocks in the sector. In the water sector, the firm's relative capital productivity change is determined by comparing per unit capital cost forecasts on individual activities (eg, laying a pipe) across firms. In the electricity distribution sector, the regulator considers the level of historical and forecast capital investment across firms. Ofgem have not published information on the actual targets used. Ofwat's targets for capital expenditure are presented in Table A.13.

Table A.13: Capital efficiency targets in the water sector

Review	Target for period
1994	Water: 0-8.7%
	Sewerage: 0-8.5
1999	Water: 9-24%
	Sewerage: 7-19%

Source: Ofwat - various publications

In the 1999 distribution price review, Ofgem (1999c) adjusted the allowed revenue of three companies by 0.25% for apparent overforecasting. This contrasts with our assumption that the regulator does not make ex-ante adjustments for expected forecasting bias. This is the only time that a regulator has explicitly done this, however, and it will be interesting to see whether it becomes a more prominent element in the price cap setting process in future reviews.

Finally, both the water and electricity distribution regulators have considered making output-delivery adjustments to the investment forecasts. In the 1994 distribution review Offer (1993) considered whether the investment programme assumed for the 1991 to 1995 period had been carried out. It was argued that this retrospective review was needed to 'ensure that further capital expenditures are justified on the basis of a rigorous analysis taking into account compliance with statutory obligations and planning standards, the effects on quality of service, and the likely costs and benefits involved'. In the 1994 final proposals no reference was made to any adjustments made to reflect unjustified underspend relative to that which was allowed at privatisation. In the 1994 water review Ofwat (1993c) examined the firm's record at delivering service to customers and proposed that 'offsetting adjustments will be made to costs or capital values, so that customers do not pay for standards which they have not received'.

We assume in the RPI-X game that the regulator takes the outputs as given when calculating efficient net new investment. In practice a number of iterations usually take place here with the economic regulator, and other regulators, testing the impact of changes in the output targets on required costs. This only occurs if the targets are considered flexible in any way.

Cost of capital

We have assumed that the regulator calculates a single value for the cost of capital using historical financial market information. In practice, a range of cost of capital values are usually calculated. Table A.14 provides examples of the ranges which were used in recent reviews. There has not been much variation in the value used across the sectors, despite the lengthy and detailed debates which arise at each review in relation to this variable[16].

Most regulators have used the capital asset pricing model (CAPM) to calculate the allowed cost of capital level[17]. Under this model the cost of capital is a weighted sum of the required return on equity and the required return on debt financing. That is:

$$\begin{aligned} r &= (1-g)COE + (g)COD \\ &= (1-g)\left[r^* + (ERP \times \beta)\right] + gY \end{aligned}$$

[16] A small change in the cost of capital has a significant impact on the firm's revenues, given the size of the asset bases in the utility sectors. This is why this variable is so prominent in discussions at all periodic reviews.

[17] This is confirmed by Buckland and Fraser (2001).

Here r is the cost of capital, COE is the cost of equity, g is the level of gearing and COD is the cost of debt. The cost of equity is equal to the return earned on a riskless asset - the risk-free interest rate r^*- plus the additional expected return from a risky stock. This additional return is equal to the equity risk premium (ERP) times the firm-specific equity beta (β). The equity risk premium measures the differences between the return earned on a market index and the risk free rate. The equity beta measures the systematic riskiness of the firm relative to the market index. The cost of debt is estimated as the observable yield on the firm's debt, Y.

The regulators compare the results from the CAPM model to those of other financial models - eg, the dividend growth model - and to the cost of capital values used by other regulators and by the Commission. When deciding on the final value the regulator also consults with key players in the financial market and academic experts. In those sectors where it has become increasingly difficult to obtain financial information on the regulated firms - because of takeover and merger activity - the regulators have used comparator cost of capital estimates from other utility sectors.

Table A.14: Allowed cost of capital

Review	Range
British Gas (pre-tax)	
1992	5-7%
1993 (MMC)	4-4.5% on existing assets
	6.5-7.5% on new investments
WASCs	
Privatisation	7%
1994 (post-tax)	5-6% on new investments
	6-7% on existing assets (by 2005)
1995 (MMC, pre-tax)	6-8%
1999 (pre-tax)	4.25-5.25% (+0.75% for WOCs)
2000 (MMC, pre-tax)	5.6% (Mid Kent Water)
RECs (pre-tax)	
1994	7%
1995 (MMC)	7%
1999	6.5%
NGC (pre-tax)	
1996	7%
2000	6.25%

Source: Armstrong, Cowan and Vickers (1994), OXERA (1998a, 1999), Ofgem, Ofwat and Offer - various publications

Depreciation

We assume that the regulator sets the allowed depreciation charge equal to annualised required maintenance expenditure over the expected life of the asset. The asset lives are determined by analysing the current physical state of the network and by considering the rate at which the assets are expected to depreciate. Here the regulator relies on information provided by the firm about the physical state of the network assets.

As assumed in our description, the regulators use historical maintenance expenditure levels as the basis for setting future expenditure requirements. These are adjusted downwards for suspected inefficiency. For example, in the 1994 water review Ofwat (1994) argued that 'Assets should be properly maintained, but customers should not have to pay for gold plating'.

The water and electricity distribution regulators assess whether the firm's historical investment levels are efficient by comparing them with those of other firms in the sector. In other sectors, the regulators undertake benchmarking analysis with firms in other industries. Regulators also employ engineering consultants to undertake detailed studies of the firm's expenditure requirements. These are still based on company information about the physical assets. With these comparisons the regulator forms a judgement on what the efficient level of maintenance expenditure is.

Once the total amount of required maintenance expenditure has been forecast the regulator calculates an annual depreciation charge which will cover the net present value of the expenditure over a number of years. Table A.15 summarises the methodologies used by the regulators to calculate the depreciation charge. In general straight line current cost depreciation is used. Using this methodology the value of depreciation does not change significantly from one period to the next, except where a large proportion of assets reach the end of their lives[18].

[18]The net present value of expected maintenance expenditure is always recouped by the firm, whatever the methodology used. The timing at which funding becomes available varies with different methodologies. In this regard the firm and the regulator may have different preferences about whether the expenditure should be spread evenly over a long time period, front-loaded or recouped at the time when the assets have almost fully depreciated.

Table A.15: Calculating the depreciation charge

Review	Methodology used
RECs	
1994	Uniform over 10-15 years for flotation assets
	Uniform over 33 years for post-vesting assets
1999	Tilted over 20 years for post-vesting assets
WASCs	
1994	Average annual spend over asset life
1999	Average annual spend over 23-years

Source: Ofwat, Offer and Ofgem - various publications

Allowed revenue, service performance and the X-factor

We assume, in our description of the RPI-X game, that the regulator makes a quality of service adjustment to the level of allowed revenue in the first year of the regulatory period. This type of adjustment was introduced in the 1999 electricity distribution and water price reviews. This was the first time that a formal link was introduced between quality of service and allowed prices[19].

The initial price cut was adjusted by 0.5% to −1% in the water sector for those firms who were identified as having service performance which was significantly above or below average. The adjustments in the electricity distribution sector ranged from 0.5% to -0.375%. The size of the adjustment was chosen arbitrarily. In the water sector, the negative adjustment was applied to three WASCs and the positive adjustment was applied to two WASCs. The positive adjustment was applied to three companies in the electricity distribution sector, and the negative adjustment was applied to two companies. No adjustment was made for the other companies (five WASCs and seven RECs).

Ofwat are currently working on improving their quality of service index for the 2004 price review and the size of the allowed revenue adjustment may be increased. There has been a greater amount of activity in this area in the electricity sector. Ofgem announced, during the 1999 distribution review, that it would undertake further research into the appropriate way to measure quality of service, and that it would design a more robust performance adjustment during the regulatory period. This

[19]The idea of linking standards of service achieved to the price control was also raised in the 1994 distribution price review. Offer concluded that such a regime would not be feasible until consistent measures for quality of service variables were determined and data collected over a number of years. The regulator therefore focused attention on gathering the required information between 1995 and 2000 rather than designing a specific mechanism at that point.

research was undertaken in the Information and Incentives Project (IIP) between 1999 and 2001, and resulted in a formal licence change in April 2002[20].

The April 2002 licence change formalised the service performance adjustment by adding an extra 'service parameter' in the formula for the allowed change in average revenue. The parameter is negative for those companies who fail to meet quality of supply targets during the price control period. The amount of the adjustment is limited to 1.75% of revenue in each year of the period. The parameter is positive for those companies who deliver a quality of supply higher than their 2004/05 target by the end of the period. The firms who outperformed relative to their target are banded together and the top portion get 2% of regulatory revenue. All other outperforming firms receive a prorata reward for their relative performance. This reward focuses on improvement during the regulatory period rather than annual changes.

Under the new incentive scheme, service performance is to be monitored during the 2000-2005 regulatory period and the financial adjustments will be implemented in the 2004 review. In future regulatory periods, the adjustments could be implemented on an annual basis. Ofgem also intend to base targets in the next regulatory period on performance in this period.

When annual allowed revenue has been calculated the regulator determines the required X factor profile. As noted in our description of the RPI-X game, several different combinations of X-factors can deliver the same net present value of revenue and the regulator uses his *judgement* as to which one is most appropriate. Table A.16 provides example of the range of X-factors used by regulators in recent reviews. We see that the X-factor can be constant over time or vary by year[21]. In particular the X-factor in the first year of the regulatory period, the P_0 cut, may be very large.

[20]The reader is referred to the various Ofgem reports cited in the bibliography for more detail on this project. Frontier Economics (2003b) also provide a useful summary of the Information and Incentives Project.

[21]Armstrong, Cowan and Vickers (1994) note that 'In telecommunications, gas, and electricity X factors are constant throughout the regulatory period, and it appears that regulators generally prefer to have constant factors......For BAA and most water companies, however, the X or K factors are not constant. This is presumably because of the large investment programs in these two industries'.

Table A.16: X-factors

Period	Price Cap (%)
RECs	
1990-1995	RPI+0 to RPI+2.5
1995-2000[1] (A)	P_0 cut (1995): RPI-11 to RPI-17
	X (1996-2000): RPI-2
1995-2000[1] (B)	P_0 cut (1996): RPI-10 to RPI-13
	X (1997-2000): RPI-3
2000-2005	P_0 cut (2000): RPI-19 to RPI-33
	X (2001-2005): RPI-3
NGC	
1990-1993	RPI-0
1993-1997	RPI-3
1997-2001	P_0 cut (1997): RPI-20
	X(1998-2001): RPI-4
2001-2006	X (2001): RPI-0
	X (2002-2006): RPI-3
WASCs	
1990-1995	RPI+3 to RPI+7
1995-2000[2] (A)	RPI+0 to RPI+5.5
1995-2000[2] (B)	RPI+0.5 to RPI+4
2000-2005[2] (B)	RPI+0 to RPI+3
2000-2005[2] (C)	P_0 cut (2000): RPI-9.3 to RPI-19.4
	X (2001-2005): RPI-0.5 to RPI+2.1

Note: [1]Offer agreed the price controls for the 1995-2000 period in August 1994 (marked A). In March 1995 the distribution price review was reopened because the regulator argued that information learnt since August suggested that the original contract was too lenient. The revised contracts were agreed in July 1995 (marked B). The 1996 P_0 cuts were in addition to those already agreed for 1995.
[2] At privatisation the regulatory lag was set at ten years for the water sector but the regulator had the option, which was taken-up in 1995 and 2000, to revise the existing contract after a five-year period. Price caps were therefore set for the period 1990 to 2000 at privatisation (marked A) and for the period 1995 to 2005 (marked B) at the first review in 1994. In the 1999 review the regulator changed the length of the regulatory lag to five years and prices were formally set for 2000 to 2005 only (marked C).
Source: OXERA Guides; Armstrong, Cowan and Vickers (1994); Ofwat, Ofgem and Offer - various publications

Output targets

We assumed that the regulators set output targets which do not allow for a deterioration in current standards. That is, the target for the next period is set equal to the current standard of service plus a possible upward adjustment. We outline a number of cases below where the regulator has tightened the firm's quality of service targets from one period to the next.

- During the 1994 electricity distribution price review, the DG examined the standards of service which companies delivered to customers. Offer (1994) indicated that the price control was set to 'raise levels of service in a number of areas' and, more generally, that there was an expectation that existing standards would be maintained alongside the reduction in prices. The DG increased the minimum level of some of the Overall Standards and reduced the required response time for some companies for the Guaranteed Standards.

- In the 1999 electricity distribution review, the DG introduced slightly tighter targets for availability of supply and security of supply. The targets were set by 'applying a required percentage improvement over the period of the price control (2000/01 to 2004/05) to the forecast performance for 1999/00, which was derived from a 10-year linear trend in actual performance' (Ofgem, 2001f).

- Offer considered reducing the standards set for NGC in 1995 as a means of providing lower prices to customers. It was decided (Offer, 1996d) that there should be 'no general relaxation in standards', but the company could reduce some standards in some circumstances (eg, good weather) as long as 'there was no widespread loss of supply'. Offer also noted that 'customers should be offered more choice about the level of security supply they wanted'.

- In the 1994 and 1999 water reviews, Ofwat expected companies to maintain existing standards of service, and to fund any discretionary improvements through cost savings.

Ideally targets would be based on information about the level of quality that customers are willing to pay for. This principle is often recognised explicitly by the regulators. For example, in the 1994 review, Ofwat (1993b) emphasised that the costs which are included in the price cap calculation should reflect 'whether customers are prepared to pay to maintain levels of service'. The following examples show the attempts which have been made to collect data on consumer preferences.

- Offer (1993) clearly stated that the allowed investment for the next regulatory period should be guided by information on the customers' trade-off between price and quality. A MORI survey was published which found that few customers were 'willing to receive lower standards of service in exchange for lower electricity prices' (Offer, 1994). It is not clear how, if at all, this information was built into the regulator's final decision.

- In the 1999 distribution review Ofgem (1999b) provided high-level evidence on willingness-to-pay: 'there is limited willingness to pay modest sums for quality improvements'. It is not clear what the source of this information was or how robust the findings are. Ofgem (1999d) also noted that it did not have a formal system for weighting quality of supply against costs. This was considered a limitation of the methodology used, and the regulator stated that it wanted to improve the information base in the future so that it was better able to analyse the operating cost-capital expenditure-quality trade-off.

- In the 1994 water review, Ofwat (1993b) argued that not enough was done at privatisation to 'assess whether customers were willing to pay the price of the substantial improvements in standards that were required of the companies'. The regulator therefore encouraged companies to undertake customer surveys to assess willingness-to-pay for service improvements. It is not clear how, if at all, this information was used to set the price caps. It is thought that Ofwat placed little emphasis on it as it did not trust the results of the surveys used.

- In the 1999 water review, consumer groups, companies, the EA and the DWI undertook surveys to determine customer willingness-to-pay for service improvements and for changes in drinking water quality and environmental standards. Ofwat (1998f) summarised the findings of these surveys. The evidence was used to reinforce the high-level judgement that customers did not want further increases in bills, but that they would welcome improvements if these could be delivered through efficiency savings. The specific investment programmes were not determined using regional willingness-to-pay information, however, because Ofwat (1998f) believed that 'Only broad conclusions can be drawn.....since respondents were usually not presented with detailed and costed improvements that they could trade-off against specific bill reductions'.

In most cases the collected data was not transparently used to set the service targets. This is generally because of the difficulty of collecting reliable willingness-to-pay information. One particular problem is the fact that consumers are not homogeneous and it is difficult to establish a simple measure to reflect a wide range of opinions on the appropriate price-quality trade-off. In addition regulators, and the Commission, are often sceptical about the value of information gathered from surveys. For example, Ofwat (1998f) argued that 'Questions about willingness to pay tend to reveal more about customers' values than the decision they would actually make when confronted with real choices'. The final decision on the quality of service required by consumers therefore usually reflects the regulator's judgement which will be partially influenced by the survey information collected.

A.2.2 The firm's move

The firm decides whether or not to accept the regulator's contract proposal by comparing profits under that contract to the expected profits from a Commission investigation. Table A.17 provides some examples of the differences between the Commission's contract and that proposed by the regulator. These suggest that the Commission generally provides similar price caps to the regulator. It is difficult to judge whether or not the firms involved benefited from the Commission's contract, however, as the underlying assumptions, including the output targets, could be very different. For example, the firm may be required to deliver more outputs with the same level of revenues, or it may be required to deliver higher efficiency savings but be allowed earn a higher cost of capital.

Table A.17: Comparing price cap proposals

Case	Regulator	Commission
South West Water		
1995	95/96: RPI+1.5%	95/96: RPI+1%
	96/97-99/00: RPI+1%	96/97-99/00: RPI+1%
BG plc		
1997	97/98: RPI-20%	97/98: RPI-21%
	98/99-01/02: RPI-2.5%	98/99-01/02: RPI-2%
Mid Kent Water		
2000	00/01: RPI-19.7%	00/01: RPI-19.7%
	01/02-03/04: RPI+0%	01/02: RPI+4.5%
	04/05: RPI+1.6%	02/03: RPI+3.2%
		03/04-04/05: RPI+0%
Hydro-Electric		
1995	95/96-99/00: RPI-1%	95/96: RPI-0.3%
		96/97-99/00: RPI-2%
NIE (T&D)[1]		
1997	97/98: RPI-30%	97/98: RPI-25%
	98/99-01/02: RPI-2%	98/99-01/02: RPI-2%

Note: [1]This is the price cap for Northern Ireland Electricity's Transmission and Distribution business. Separate caps are set for the Power Procurement business and for the Supply business.
Source: OXERA (1998b) and Competition Commission (2000)

In practice, firms have tended to accept the regulator's contract proposal. Armstrong, Cowan and Vickers (1994) note that 'Typically firms have preferred to avoid the uncertainties of an MMC review'. Table A.18 shows the few cases where a reference has been required in the water and electricity sectors in England and Wales. The Commission also investigated the price cap for Scottish Hydro-Electric in 1995 and for Northern Ireland Electricity in 1997. The relatively few number of Commission cases may reflect the fact that regulators take account of the 'outside option' of an appeal hearing when devising the contract proposal. They do this by making assumptions which are consistent with those used by the Commission in the past, thereby reducing the probability of the Commission offering a contract which is very different from that proposed by the regulator. This is in the regulator's interest as he also faces large costs, both financial and in terms of time, if there is a Commission investigation at the end of the periodic review process.

Table A.18: Competition Commission references

Review	Firm's decision
RECs	
1994	Accepted by all
1995	Accepted by all
1999	Accepted by all
NGC	
1992	Accepted
1996	Accepted
2000	Accepted
WASCs	
1994	9 accepted
	South West Water rejected
	Postsmouth Water (WOC) rejected
1999	All accepted
	Mid-Kent Water (WOC) rejected
	Sutton & East Surrey Water (WOC) rejected

Note: Offer agreed the price controls for the 1995-2000 period in August 1994. In March 1995 the distribution price review was reopened because the regulator argued that information learnt since August suggested that the original contract was too lenient. The revised contracts were agreed in July 1995.

A.2.3 The Competition Commission Game

We assume that the Commission first decides whether the existing cap is in the public interest. In all price review cases held to-date the Commission has found that the current price limit is not in the public interest and should be changed. This is not surprising given that firms' costs have been significantly lower than assumed at each periodic review, and that output targets have changed over time. Once this decision is made the Commission considers how the cap should be changed to deal with the public interest concerns.

We assume, in this regard, that the Commission may make different assumptions to the regulator. Table A.19 highlights a number of examples where the Commission's assumptions have been different to those used by the regulator. The details behind these assumptions are provided in the published reports. Despite these differences, the final price caps have been vary similar (see Table A.17).

Table A.19: Different allowed revenue assumptions

Review	Regulator	Commission
South West Water (1995)		
NNI (1995-2005)	£900m	£950-990m
Operating costs	£908m	£930-950m
Sharing operating savings	10-year glidepath	5-year glidepath
BG plc (1997)		
Cost of capital	5.92-7.6%	7%
RCV	£12.4bn	£11.6bn
NNI (1997-2002)	£2.4bn	£4.1bn
Mid Kent Water (2000)		
Cost of capital	5.5%	5.6%
Operating efficiency target	4.3% p.a.	3.7% p.a.
Hydro-Electric (1995)		
NNI (1995-2000)	£255.4m	£277.9m
NIE (1997)		
Operating costs	£301.8m	£328.1m

Source: OXERA (1998b) and Competition Commission (2000)

We include the firm's option to seek a judicial review in our description of the RPI-X game. In general the regulators have implemented final licence changes which closely reflect the recommendations of the Commission. In the case of water, the recommendations are implemented exactly. There has therefore only been two judicial review cases to-date, one of which was taken by a firm not involved in the original Commission investigation.

1. *ScottishPower*: on the basis of a recommendation made by the Commission, Offer adjusted Hydro-Electric's supply price control formula so that the Great Britain Yardstick was used as the basis for determining generation costs in the control. ScottishPower, the other operator in Scotland, requested that the same change be made to its price control. Offer rejected this request and the case was referred to the High Court. Offer won the judicial review case. The matter was then referred to the Court of Appeal which ruled in favour of ScottishPower in February 1997. The regulator considered making a further appeal to the House of Lords, but finally made the required amendment to ScottishPower's licence, recognising that a further appeal would be lengthy and costly.

2. *Northern Ireland Electricity*: following the Commission's report, Ofreg introduced a revised licence condition for NIE's transmission and distribution business in 1997. The details of the licence change were different to those proposed

by the Commission. In particular, the regulator made different assumptions on the size of the regulatory capital value at privatisation, on the treatment of depreciation, and on the allowed levels of capital and operating costs. When the regulator continued to refuse to implement the Commission's exact findings NIE pursued the matter via a judicial review. The High Court in Belfast supported the regulator's stance in June 1998, but the company then sought a further investigation by the Court of Appeal. In October 1998 the Northern Ireland Court of Appeal ruled in NIE's favour and the firm's licence was changed.

In both cases the firm has won the right to have the Commission's recommendations, rather than the regulator's alternative contract, implemented into its licence. The cases took a long time to resolve, however, and the outcome was shortly followed by another price review which changed the contracts again.

A.3 The implementation game

We are not able to provide evidence on how the regulated firms make their decisions in the implementation game. The decision-making processes in the boardroom are not publicly available and, hence, we rely on the standard economic assumption of profit-maximisation for our description.

We do, however, have evidence on the extent to which the regulators choose to change the regulatory contract mid-period[22]. The unexpected interventions are often driven by pressure from the government or consumer groups to reduce the profits earned by the firm. In other cases, the intervention was signalled at the periodic review - for example with cost past-through mechanisms - and was therefore expected by the firm.

- *Formal revision in response to high profits:* In the electricity distribution sector the regulator set price limits in 1994 for the period 1995 to 2000. After the final price limits had been agreed the regulator learnt that higher profits and lower costs were feasible, and he announced that the price limits were to be revised. New contracts were proposed in July 1995. This intervention involved a complete and formal revision of the regulatory contract - essentially a second periodic review.

[22]The extent of mid-period intervention in the mid- to late-1990s is usefully summarised by Williamson (2001) as follows: 'A windfall tax has been levied, and regulators have reopened reviews, clawed-back gains from outperforming revenue caps and revisited revenue blocks such as the regulatory asset base and depreciation'.

- *Formal revision in response to changing market conditions:* British Gas's regulatory contract was formally revised in 1993 when the MMC recommended a relaxation in the price cap for transportation. This was proposed as compensation for the introduction of competition in the small customer market at a faster pace than expected. Similarly, during the 1991 duopoly review in the telecoms sector, British Telecom's price cap was changed from 4.5% to 6.25% for the last two years of the existing regulatory period (1992 to 1994). These are examples of how a change in policy - in this case market structure considerations - can lead to a complete reopening of the regulatory contract.

- *Formal revision in response to changing input prices:* In 1992 and 1993 the construction price index fell significantly, reducing the capital investment costs of water companies below what was assumed when prices were set at privatisation. In response to this Ofwat introduced a licence change to revise the price cap of seventeen companies downwards in 1993 and 1994. This is an example of a partial formal revision to the contract.

- *Formal revision in response to tighter output targets:* Upward revisions to the price cap have also been observed in the water sector when output targets have been changed during the regulatory period. Whether or not the contract is formally revised depends on the cost of delivering the new targets. If the regulator believes that the efficient costs can be covered under the existing price limit the contract is not revised. The regulator will, however, include any additional capital investment in the regulatory capital value at the next periodic review. If, however, the financial cost of the required output is large, and the firm is unable to 'finance its functions' under the current price limit the regulator will, usually at the firm's request, hold an interim price review. This results in a formal and full revision to the contract mid-period.

- *Informal revision in response to high profits:* Shortly after the new price limits came into effect in 1995 the WASCs made large profits and dividend payments. In response, the regulator recommended that companies deliver the benefits of cost savings to customers immediately by charging prices which were below those allowed under the price limit[23]. That is, the firms were encouraged to operate as if an implicit lower price cap was set. This form of benefit sharing is an example of an informal adjustment to the 'spirit' of the regulatory contract, rather than an actual revision to the detail of the written contract.

[23]Although no formal mechanism was used to implement such benefit sharing schemes, and no indication was given of what would happen companies who didn't adopt this approach, several companies voluntarily set lower prices in the latter years of the 1995-2000 regulatory period.

- *Automatic pass-through of costs:* In the gas, electricity supply and airports sectors provisions are in place to allow some costs to be passed-onto customers as soon as they are incurred[24]. This results in an automatic adjustment to the price cap during the regulatory period. As this is built into the regime it does not involve any revision of the contract. The types of costs which firms are allowed passthrough are as follows - (1) Electricity supply companies were able to pass on wholesale prices, transmission charges, distribution charges and the fossil fuel levy to customers. (2) British Gas is able to pass through an index of gas purchase costs. Between 1987 and 1992 all purchase costs we passed through. Since 1992 the company has only been able to pass through a gas price index less 1%. (3) BAA can pass through up to 95% of security costs in its airport charges.

The regulator's decision to change the contract is made in reaction to one of three events - a change in costs because of exogenous shocks, an increase in costs because of the introduction of new outputs, and a higher than expected improvement in the firm's efficiency level. In general, the non-economic regulators have attempted to coordinate the timing of output decisions with the periodic review so that changes mid-period are minimised. The legislative timetable does not follow the regulatory timetable, however, and changes may therefore arise at any time. This is particularly notable in the water sector where output targets are driven by EU Directives and, as noted by Ofwat (1998f), 'History has shown that it is likely that further new obligations will be imposed before the next review'. We should therefore expect intervention in response to output changes to continue. The concern, however, is whether regulators will continue to change the contract in response to the high-profits which firms earn *because* the RPI-X mechanism is working as expected.

[24] A theoretical analysis of the implication of cost pass-through for allocative efficiency and technical efficiency can be found in Holzleitner (2001). We do not discuss this issue here.

Appendix B

The regulated businesses

We calculate productivity growth rates for the water, electricity transmission, and electricity distribution sectors in Chapter 4. The analysis is based on data for the following companies:

- the ten *water and sewerage companies* **(WASCs)** - companies which hold a licence to provide monopoly water and sewerage services within a specified region in England and Wales;
- the *distribution business of the twelve regional electricity companies* **(RECs)** in England and Wales - companies which hold a monopoly licence to distribute electricity from the high-voltage network, along lower-voltage power lines, to the end-user within a specified franchise area[1]; and
- the *transmission business of the National Grid Company* **(NGC)** - the company which holds the monopoly licence to transmit electricity from upstream generation plants along high-voltage lines to the distribution network. The licence area covers all of England and Wales.

Other companies of interest, which could be included in an extended version of this analysis, are BT (formerly British Telecom), Transco (formerly British Gas) and Railtrack. BT and Transco have been excluded as it is difficult to collect a consistent dataset which is focused on the regulated network monopoly elements of these businesses. Railtrack has been excluded as the available dataset is relatively short.

The companies considered here undertake a number of different activities in addition to the provision of their regulated services. We wish to clarify precisely what

[1]These twelve companies are also referred to as Public Electricity Suppliers (PESs), along with ScottishPower, Scottish Hydro-Electric and Northern Ireland Electricity.

element of each company's activities are of interest to us to ensure that the appropriate data is abstracted from published regulatory accounts. The regulated businesses which we focus on are those which the regulator applies a stand-alone price cap to.

- *WASCs* - the WASCs were privatised on December 31st 1989 and floated on the London Stock Exchange on December 11th 1990. The regulator sets a single price cap (the K factor) for the provision of water and sewerage services. That is, the combined provision of water and sewerage services is treated as a single network monopoly by the regulator. We also adopt this approach, while recognising that there may be scope to introduce competition in some elements of the business (eg, abstraction or customer services)[2]. We also note that, by considering both businesses together, and by combining all elements of service provision (eg, treatment, distribution, customer services), we are unable to consider important efficiency questions relating to the individual elements of the production process.

- *Electricity distribution* - the RECs were privatised on March 31st 1990 and floated on the London Stock Exchange on December 11th 1990. They are responsible for the supply and distribution of electricity. Initially both activities were regulated, but over time competition has been introduced into supply. Distribution remains the core network monopoly activity of these companies. We therefore focus attention on the distribution business of these companies only. The regulator has always set a separate price cap for the distribution business.

- *Electricity transmission* - the National Grid Company was privatised on March 31st 1990. It was initially owned jointly by the RECs but was floated on the stock exchange on December 11th 1995. When setting the price cap for NGC's activities, the regulator considers the revenues earned from transmission services, settlements, ancillary services and interconnectors. These are the activities which we also focus on in our productivity analysis.

There are three business units associated with each of the regulated firms. Each of these publish annual accounting information.

1. The *regulated business* which operates under the price cap publishes annual regulatory accounts. The companies are required to submit these accounts to the regulator as part of their licence conditions. The water regulator provides Regulatory Accounting Guidelines (RAGs), ensuring consistency across the companies. No guidance has been provided by the electricity regulator, however, and

[2]See Robinson (2000) for a discussion of the prospects for competition in the water sector.

hence there may be some concern about consistency in the accounting methodology used by companies - for example with respect to allocating common costs across the supply and distribution businesses. Concerns have been expressed by the electricity regulator in this regard and work is ongoing to develop a set of accounting guidelines for the sector.

2. The *Group,* which owns the regulated business and is the holder of the monopoly licence, is a statutory company which publishes its own annual report and accounts under the Companies Act 1985. This company undertakes activities in addition to the running of the regulated business.

3. The *Parent* company, which owns the Group business and is listed on the Stock Exchange (but potentially not in the UK), is also a statutory company which publishes its own annual report and accounts. In many instances, the regulated business is only a small subsidiary of the parent company.

The Government retained a golden share in the water and regional electricity companies until March 1995. After this date takeover activity began and most of the companies in this analysis have been owned by more than one parent company since. Tables B.1 and B.2 provide details of the regulated companies, outlining the name of the licence-holder and all its parent companies. NGC has not been subject to any takeover activity and remains the only transmission business in England and Wales. Stock market information, and data in the accounts of the Parent and Group companies, relate to businesses which encompass many activities. This information is therefore a misleading indicator of the regulated business's financial position.

We source all financial data from the regulatory accounts which are published using the current cost accounting methodology. These accounts provide information on each of the service areas in the regulated business. For example, separate accounting data is provided for the supply and distribution businesses of the RECs, and separate data is available for water services and sewerage services for the WASCs. All non-financial (physical) data is taken from documents published by the regulatory agencies or by the Centre for Regulated Industries (CRI).

Table B.1: Ownership of the electricity companies to March 31st 2000

Regulated firm	Group	Parent
Distribution		
Eastern	Eastern Group	TXU Europe
		The Energy Group: 02/97-09/98
		Hanson: 09/95-02-97
		Group: 03/90-09/95
East Midlands	East Midlands Electricity plc	PowerGen (f)
		Dominion Resources: 01/97-07/98
		Group: 03/90-01/97
London	London Electricity plc	EDF
		Entergy Corporation: 07/97-12/98
		Group: 03/90-07/97
Manweb	Manweb plc	ScottishPower (f)
		Group: 03/90-10/95
Midlands	Midlands Electricity plc	GPU Power UK
		Avon Energy Partners: 06/96-06/99
		Group: 03/90-06/96
Northern	Northern Electric plc	Berkshire Hathway Group
		CE Electric UK: 12/96-03/00
		Group: 03/90-12/96
Norweb	Norweb plc	United Utilities (f)
		North West Water plc: 11/95-04/96
		Group: 03/90-11/95
Seeboard	SEEBOARD plc	American Electric Power Co
		Central & Southern: 01/96-12/97
		Group: 03/90-01/96
Southern	Southern Electric plc	Scottish and Southern Energy
		Group: 03/90-12/98
Swalec	SWALEC plc	Western Power Distribution
		Hyder: 03/96 to 09/2000
		Welsh Water plc: 01/96-03/96
		Group: 03/90-01/96
Sweb	SWEB plc	Western Power Distribution
		Southern Company: 09/95-01/99
		Group: 03/90-09/95
Yorkshire	Yorkshire Electriciy plc	American Electric
		Group: 03/90-04/ 97
Transmission		
NGC	National Grid Group	National Grid Group (f)
		Jointly owned by RECs: 03/90-12/95

Table B.2: Ownership of the water companies to March 31st 2000

Regulated firm	Group	Parent
Anglian	Anglian Water plc	AWG plc (f)
Dwr Cymru	Welsh Water plc	Glas Cymru Cyfyngedig
		WPD: 09/00-05/01
		Hyder plc: 01/96-09/00
		Welsh Water plc: 12/89-01/96
Northumbrian	Northumbrian Water Group plc	Lyonnaise des Eaux
		Group: 12/89-02/96
North West	North West Water plc	United Utilities plc (f)
Severn Trent	Severn Trent plc	Severn Trent plc (f)
South West	South West Water plc	Pennon Group plc (f)
Southern	Southern Water plc	ScottishPower plc: 07/96-?/02
		Group: 12/89-07/96
Thames	Thames Water plc	RWE AG
		Group: 12/89-11/00
Wessex	Wessex Water plc	Enron Corp: 09/98-?/02
		Group: 12/89-09/98
Yorkshire	Yorkshire Water plc	Kelda Group plc (f)

Appendix C

Productivity variables

To measure productivity we need to define the outputs produced, and the inputs used to produce these outputs. We consider the most suitable *available* measures for each of these variables here.

C.1 Output

Each regulated business is required to ensure that supply meets customer demand at all times[1]. In addition, the firms are expected to provide a high quality of supply to consumers. As such, there are essentially two distinct but related outputs produced - the physical output and the quality of supply provided. We wish to capture both outputs in our productivity measures, particularly as the investment programmes undertaken since privatisation have been aimed at improving the quality of supply, rather than simply increasing capacity to provide extra volume to customers[2].

C.1.1 Output level

The amount of the physical output produced is, essentially, the volume of the good which is transported along the network. The specific measures used for each sector are as follows.

[1]In the water sector, the regulated firm is involved with all stages of production and the final product provided is the delivery of water and sewerage services to customers. The electricity distribution companies and NGC are responsible for providing network access, and transportation services, between upstream and downstream providers. The products which are regulated are thus different - intermediate and final services - but the form of regulation is the same across all the sectors.

[2]This point is also emphasised by Saal and Parker (2001). They note that in the water sector 'a substantial portion of the additional capital input has been concerned with water quality enhancement, as well as capital maintenance. Base water and sewerage output, by contrast, has been fairly static'. Inputs are therefore used to produce an increasing quality level rather than an increase in volumes delivered.

- For NGC we measure the level of output as the volume of electricity sales to domestic, commercial, industrial and other customers. The annual data is taken from the company's Seven Year Statement (January 2001).

- The level of output for the electricity distribution companies is measured as the units of electricity distributed in the company's franchise area. The data is taken from the Centre for Regulated Industries's annual report, *UK Electricity Industry Financial and Operating Review*, for the years 1991 to 1999. The corresponding data for 2000 is taken from the Electricity Association's report, *Electricity Industry Review*.

- There are problems using a volume measure as the output variable in the water sector. A large proportion of customers receive an unmetered water supply, and the amount of water consumed by end-users is estimated by the companies[3]. The data on the volume of sewage collected is also estimated. Measured data is available on the volume of water which is put into the distribution system (input), but this does not necessarily correspond to the amount that is delivered because of leakage[4]. There is thus no reliable data on volumes delivered. We therefore use the number of properties connected to the network as a proxy measure for the amount of output produced in the water and sewerage sector[5]. This variable is calculated as the sum of the properties connected to the water system and the properties connected to the sewerage system. Data for the period 1993 to 2000 is taken from the companies' July Returns submissions to the regulator. The data for 1991 and 1992 is taken from the OFWAT annual report, *Report on Levels of Service for the water industry in England and Wales* (1991 and 1992). Data is only available in 1991 for water properties. We therefore estimate the 1991 values for sewerage propeties by assumming that the company-specific growth rate in sewerage properties from 1991 to 1992 was the same as that observed in 1992 to 1993. .

Figure C.1 shows the output index for each of the sectors for the period 1991 to 2000. Output has increased at a steady rate in all sectors, with higher growth in the

[3]In 2000 only 16% of all water customers- household and non-household - were metered.

[4]Leakage from the system has a significant effect on changes in the volume of water put into the system. Companies have made large investments to reduce leakage in recent years. This could be construed as a reduction in the level of output, even though the amount consumed may not have changed. This was noted by Waddams Price (2000) who emphasised that improvements in the leakage record of the firms will lead to a reduction in the volume measure at a time when consumption is known to have been increasing.

[5]The number of customers would be an alternative proxy for the output variable. These numbers are also estimated by the firms however and show a high degree of variability, particularly for sewerage services.

electricity sector than the water sector.

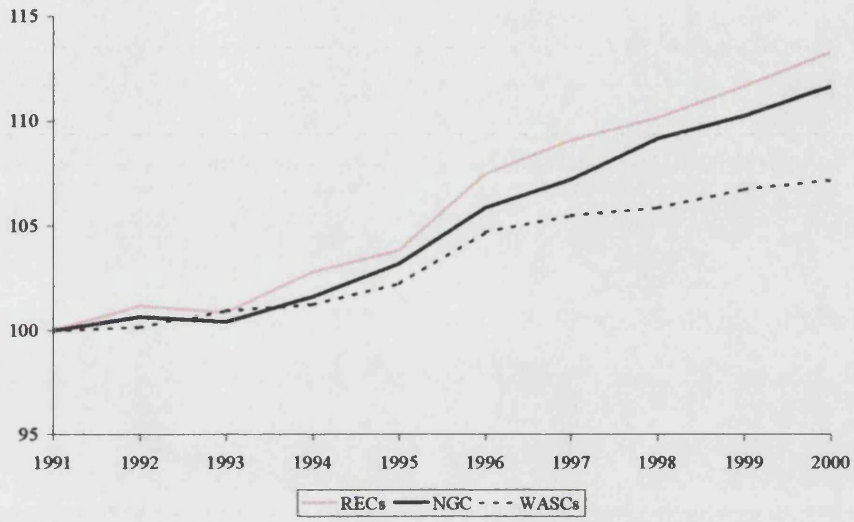

Figure C.1: Output Index

C.1.2 Quality of supply

In addition to ensuring that consumer demand is satisfied, the regulated firms are required to provide a sufficient quality of supply. From the perspective of productivity measurement, this means that the firm's inputs are used to deliver both increased volumes and quality of supply improvements. If the quality element is not captured in the productivity growth rate, the firm will be considered to have low productivity improvement even when it is producing more with a given set of inputs. This is because the additional quality being produced - 'more output'- is not reflected in the standard measure. We attempt to overcome this shortcoming by using a quality-adjusted output variable in our calculations.

In the electricity sector quality of supply is primarily focused on the need to ensure that customers get a continuous supply of electricity (ie, security of supply). The measure of interest for electricity distribution is the number of supply interruptions per 100 connected customers. The data is taken from OFGEM (2001) *Report on Distribution and Transmission Performance 2001.* The quality of supply variable used for NGC is Average Transmission System Unavailability. The data for 1992 to 2000 is taken from the OFGEM (2001) *Report on Distribution and Transmission Performance 2001.* The data for 1991 is taken from OFGEM (2000b), *The Transmission Price Control Review of the National Grid Company from 2001: Initial Thoughts Consultation Document.*

Quality of supply in the water industry is more complicated as there are two aspects to quality for each service provided. First, the quality of the service provided to customers matters. For water services we use two measures of quality of service - the percentage of properties with unplanned supply interruptions greater than 12 hours, and the percentage of properties with water pressure levels below the reference set by the regulator. The quality of service measure used for sewerage services is the percentage of properties at risk of flooding from sewers (once in ten years and twice in ten years). Data on these measures is taken from the OFWAT annual report, *Report on Levels of Service for the water industry in England and Wales* (various years).

The second aspect of quality of supply which is important in the water industry is environmental quality and drinking water quality. The drinking water quality measure used is the percentage of tests carried out which comply with the prescribed concentration value[6]. This data is taken from the Drinking Water Inspectorate's annual report, *Drinking Water Report by the Chief Inspector* (various years). For sewerage services, the impact of the disposal of sewage on the natural water environment is of most importance. This is regulated by providing each company with discharge consents which set a limit on the amount of chemicals which can be included in the discharged sewage. The quality measure used here is the percentage of sewerage treatment works complying with numeric discharge consents. The data is taken from the Government's publication, *The Digest of Environmental Statistics* (various years)[7].

We construct a quality-adjusted output measure using these quality of supply variables. The measure captures the fact that the firm produces two products - the *level of output* and *the* ***level*** *of quality of supply.* Specifically, the combined measure reflects the proportion of output which is of 'good quality' in each year. An increase in this variable indicates that the firm is producing more for consumers and we assume that this increase is welfare-improving (ie, desirable).

The quality-adjusted output index is calculated for electricity distribution and NGC as:

$$QVol_t = Vol_t \times GQ_t$$

Vol_t is an index of the level of output in the sector (1991=100). GQ_t measures the proportion of output which was of good quality in year t. For the distribution companies this is the percentage of customers who did not suffer a supply interruption. For NGC this is the percentage of the system which was available during the year.

[6]This is the numerical value assigned to water quality standards defining the maximum legal concentration.

[7]The Digest of Environmental Statistics was published by the Department of the Environment (1991-1997), the Department of Environment, Transport and the Regions (1997-2001), and the Department of Environment, Food and Rural Affairs (2001).

A similar measure is used for the water sector but it is complicated by the presence of numerous quality measures. The quality-adjusted output index is:

$$QVol_t = (R_t^w \times Vol_t^w \times GQ_t^w) + (R_t^s \times Vol_t^s \times GQ_t^s)$$

R_t^w is the share of total revenue from water services, Vol_t^w is an index of the level of output in the water sector (1991=100) and GQ_t^w is the proportion of output which is of good quality. The variables with 's' superscripts provide the corresponding information for sewerage services. The weighted measure is intended to represent the proportion of water supplied and sewage collected which is of overall 'good quality' from the customer's perspective.

Good quality in the water sector, GQ_t^w, is measured as a weighted sum of the water quality measures. Equal weights have been applied to each measure given the absence of any hard information on willingness to pay for each aspect of quality of supply. This gives us:

$$GQ_t^w = \frac{DW_t}{3} + \frac{WP_t}{3} + \frac{WI_t}{3}$$

where:

- DW_t is the proportion of the firm's tests complying with drinking water quality standards in year t;
- WP_t is the proportion of properties which *have not had* water pressure below the reference level in year t; and
- WI_t is the proportion of the firm's properties which *have not had* unplanned supply interruptions in year t.

Good quality in the sewerage sector, GQ_t^s, is measured as a weighted sum of the sewerage quality measures. Equal weights have again been applied to each measure. This gives us:

$$GQ_t^s = \frac{SF_t}{2} + \frac{DC_t}{2}$$

where:

- SF_t is the proportion of the firm's properties which are *not at risk* of sewer flooding once or twice in ten years; and
- DC_t is the proportion of the firm's discharge consents complying with the standards set

Figure C.2 shows the quality-adjusted output index for each of the sectors. There has been a significant improvement in the proportion of output which is of 'good quality' since privatisation. This output measure has increased by slightly more than the standard output index between 1991 and 2000 and there is more variation in this index from one year to the next. For example, NGC's index declined slightly in 1992 and increased significantly since then. This reflects the fact that tranmission system availability fell from 92% to 91% in 1992 and increased annually from then, reaching 96% by 1997 and remaining at that level up to 2000.

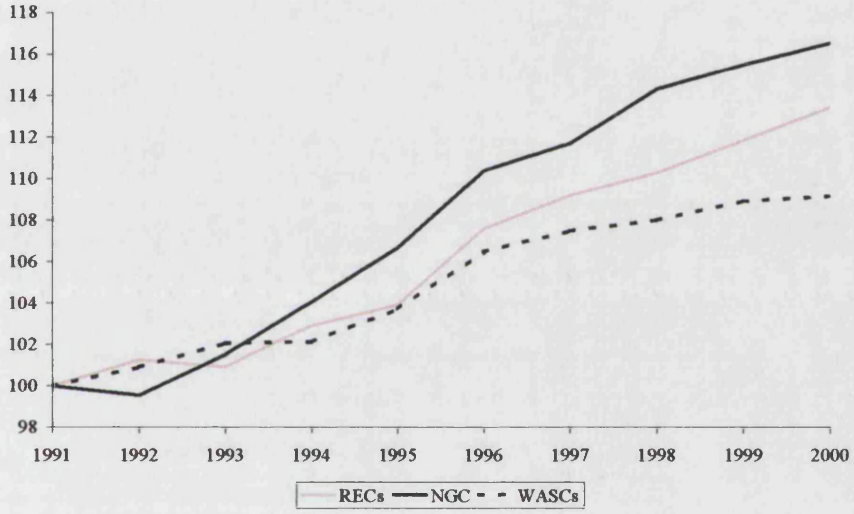

Figure C.2: Quality-adjusted Output Index

We could, alternatively, consider the second quality product as being the **size of the *improvement*** *in quality of supply* over time, and construct a combined measure using this variable. This approach was adopted by Saal and Parker (2001) and Hunt and Lynk (1995). The index used in this case would be:

$$QVol_t = Vol_t \times QI_t$$

where:

- $QVol_t$ is the quality-improvement adjusted volume in year t;
- Vol_t is an index of the level of output (with 1991=100); and
- QI_t is a quality improvement index. The quality improvement index is calculated as:

$$QI_t = \frac{Q_t}{Q_{1991}}$$

where Q_t is the individual firm's level of quality in year t and $\overline{Q}_{1991}$ is the average level of quality in the sector in 1991. For $QI_t > 1$ we see that the quality of supply for the firm is greater than the average quality in 1991 (ie, there has been a service improvement). A value of $QI_t < 1$ indicates a deterioration in quality relative to the sector average in 1991.

We found that, with our data, the TFP growth rates are biased by outliers with this methodology. In addition, the level of quality, rather than the size of the improvement in the level of quality, is thought to better reflect the actual amount produced using the given inputs in each year. We also question whether it is an appropriate measure of 'desirable output' as it seems to suggest that only increasing rates of improvement are required by consumers. In practice we would expect that high quality levels are demanded and, once these are achieved, there is limited value from increased improvements. Indeed, it is unlikely that large improvements can be continuously delivered in all years.

C.1.3 Value of output

The value of the output produced is the turnover (£m) of the regulated business. This data is obtained from the firms' annual regulatory accounts and converted to 2000 prices using the ONS Producer Price Index[8]. This data was taken from ONS, *Economic Trends*, various years. The firm's turnover reflects the amount produced and the amount charged for each unit of output. It is therefore affected by price cap regulation which restrains the level of charges over time.

Figure C.3 shows the level of turnover in each of the sectors since privatisation. The trend reflects changes in the price cap allowed (see Table A.16). In particular, the price cap has been positive in the water sector (RPI+K) and negative, on average, in the electricity distribution sector. NGC's price cap was non-negative for most of the period.

[8] In NGC's regulatory accounts data for 1991 to 1994 was classified into the following four categories: electricity transmission, ancillary services, interconnectors plus generation and settlements. From 1995 to 2000 the categories were electricity transmission, ancillary service, interconnectors and settlements. The 1995 accounts also presented information on interconnectors only for 1994. This allowed us to determine the proportion of interconnectors plus generation activity which was interconnectors only. To ensure the time series was consistent we removed data on generation activities from the 1991 to 1993 data. This was done by applying the proportion of the interconnector plus generation value which was interconnector only in 1994 to the aggregate data in the 1991 to 1993 accounts. In this way all information relating to generation activities pre-1995 was removed. **This adjustment was made for all financial data taken from NGC's regulatory accounts.**

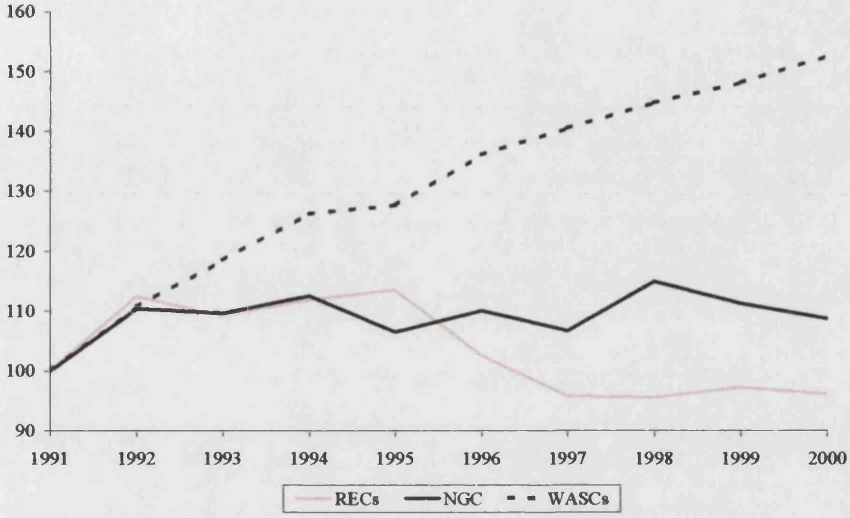

Figure C.3: Turnover Index

C.2 Inputs

We assume there are two input categories in the production process - capital and non-capital. The data used for each are discussed here.

C.2.1 Capital

Capital level

The level of capital for the regulated firm, most of which is sunk, is defined as the current cost value of gross fixed assets (£m) at the year-end[9]. This data is taken from each firm's annual regulatory accounts. The data has been subject to a number of revaluations over time. These revaluations reflect improvements in information about the network (notably MEA revaluations in the water sector), and changes in the way in which historical cost data is converted to its current cost counterpart. These changes result in a significant degree of variability in the numbers, without there necessarily being an underlying change in the level of capital available for the production process. We therefore remove the revaluations to create a consistent series for the current cost value of the fixed assets. Saal and Parker (2001) use a similar approach. They argue that this is consistent with the 'inventory' method of calculating capital values as the change in gross fixed assets now simply represents the level of net investment. That

[9]This variable is also used by Tilley and Weyman Jones (1999) in their analysis of total factor productivity growth in the electricity distribution sector.

is:

$$GA_t = GA_{t-1} + I_t - D_t$$

where GA is gross fixed assets, I is investment (ie, additions to the capital base) and D is disposals (ie, reduction in the capital base). All data has been converted to 2000 prices using the ONS Gross Fixed Capital Formation Price Deflator. This data was sourced from the ONS, *UK National Accounts* (the Blue Book), various years.

Figure C.4 shows the change in the capital level over time for each of the sectors. The capital level appears to change cyclically and the cycles coincide with the regulatory periods. The capital level in the water and electricity distribution sectors has increased between 1991 and 2000. In the electricity transmission sector, however, it was close to its 1991 level by the end of the period.

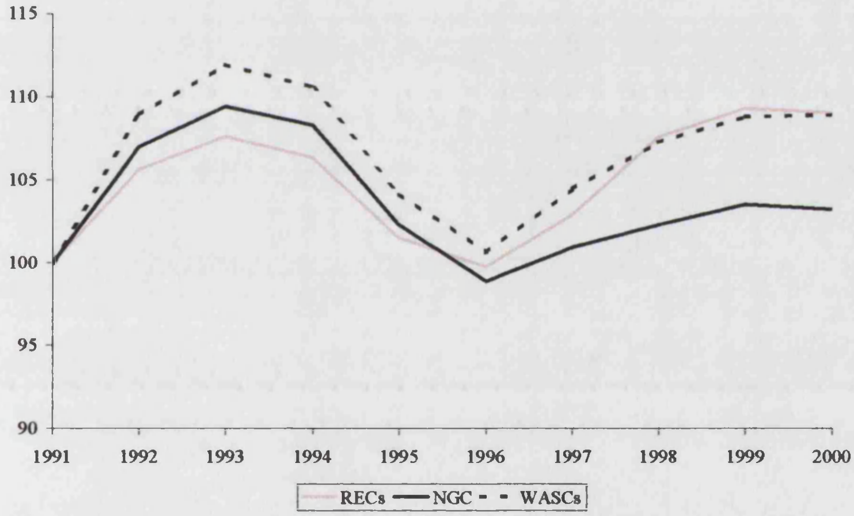

Figure C.4: Capital Level Index

Capital costs

We calculate the annual capital cost as the sum of the financing cost of capital plus the cost of depreciation:

$$(\text{Rental rate} \times \text{Capital level}) + \text{capital charges}$$

The average real discount rate on a treasury bill is used for the rental rate of capital. This data is taken from ONS, *Economic Trends*, various Years. In the electricity sector the capital charge is simply the annual depreciation charge. In the water sector, infrastructure assets (those below ground) are financed through an infrastructure renewals charge and non-infrastructure assets are financed through a standard depreciation charge. The sum of these annual charges is used here. Data on capital charges is taken from the firms' annual regulatory accounts and converted to 2000 prices using the ONS Gross Fixed Capital Formation Price Deflator.

Figure C.5 shows the capital costs for each sector. The capital costs have fallen from their 1991 level. In addition, like the capital level, the costs vary cyclically and the cycles coincide with changes in the Treasury bill over time .

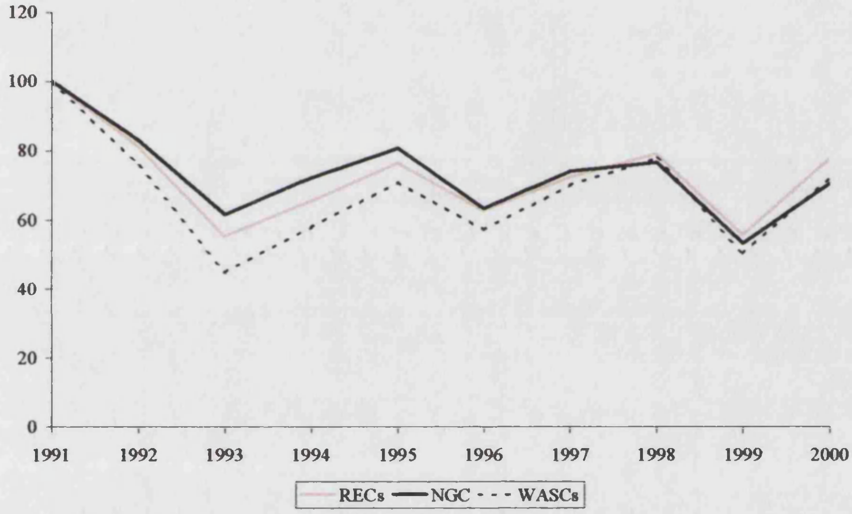

Figure C.5: Capital Cost Index

We also consider the level of gross capital expenditure when examining trends in unit costs. Gross capital expenditure is equal to the annual additions to fixed assets. The data is taken from the notes to the balance sheet in the regulatory accounts. The data has been converted to 2000 prices using the ONS Gross Fixed Capital Formation Price Deflator. Figure C.6 shows the level of capital expenditure in each of the sectors since privatisation. The level has decreased in the electricity transmission sector. Tere has been an increase in both the water and electricity distribution sectors, although expenditure has reduced in the distribution sector and has levelled off in the water sector since 1998. There is some cyclical behaviour in the level of gross capital expenditure and, again, this corresponds with the regulatory periods, perhaps with a lag.

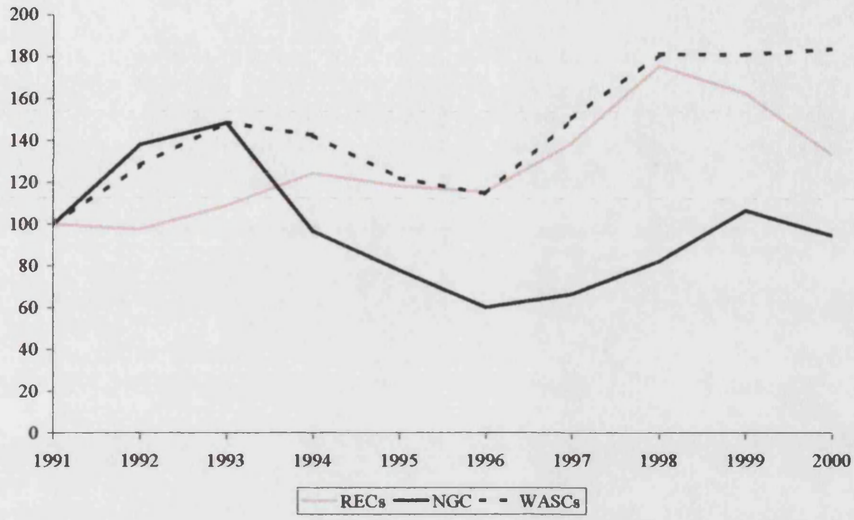

Figure C.6: Capital Expenditure Index

C.2.2 Non-capital

The level, and value, of non-capital inputs is measured as the value of the firm's controllable operating costs (£m). The financial measure is used because there is no clear physical measure which accurately encompasses all the non-capital inputs. The data is taken from each firm's annual regulatory accounts[10]. The definition of controllable operating costs varies by sector, reflecting differences in the way in which each company presents data on operating costs in its accounts. In all cases total operating costs is calculated as turnover minus current cost operating profit. Controllable operating costs is then calculated as follows.

- For NGC, controllable operating costs is equal to operating costs minus depreciation, rates, transmission services scheme direct costs, research & development costs, purchases of electricity and other costs. It is assumed that all the deducted costs are not used in the production process directly.
- In the electricity distribution sector, controllable operating costs is calculated as operating costs minus depreciation, NGC charges, research & development costs,

[10]A number of water companies did not provide a detailed breakdown of direct operating costs in their 1992/93 regulatory accounts. Information on employment costs was provided by Ofwat (by email) to ensure there was no break in the time series. This data was consistent with that sourced from regulatory accounts for all other years and is considered an accurate reflection of operating costs in 1992/93. In addition, a number of companies provided the required breakdown for 1992/93 in their 1993/94 regulatory accounts and for these companies the data coincided with that provided by Ofwat.

exceptional costs, auditors remuneration and other costs (including administrative expenses)[11]. Again, we assume that these deducted costs are not inputs to the production process.

- The water companies provide a more detailed breakdown of their direct operating costs (ie, those used in production). We are able to calculate controllable operating costs as the sum of employment costs, power costs, materials costs and the cost of hired and contracted services. There was a significant step-change in employment costs *for all companies* in 1992/1993, resulting in a structural change in the level of controllable operating costs at that point.

All the data has been converted to 2000 prices using the ONS Producer Price Index. Figure C.7 shows the change in the level of non-capital costs over time. In all sectors there has been a reduction in controllable operating costs between 1991 and 2000. The trend has been steadiest in the electricity distribution sector and most marked in the electricity transmission sector. There was a step-change in the water sector in 1993, driven by a sharp fall in employment costs, which was followed by small annual decreases up to 1998 and small annual increases between 1998 and 2000.

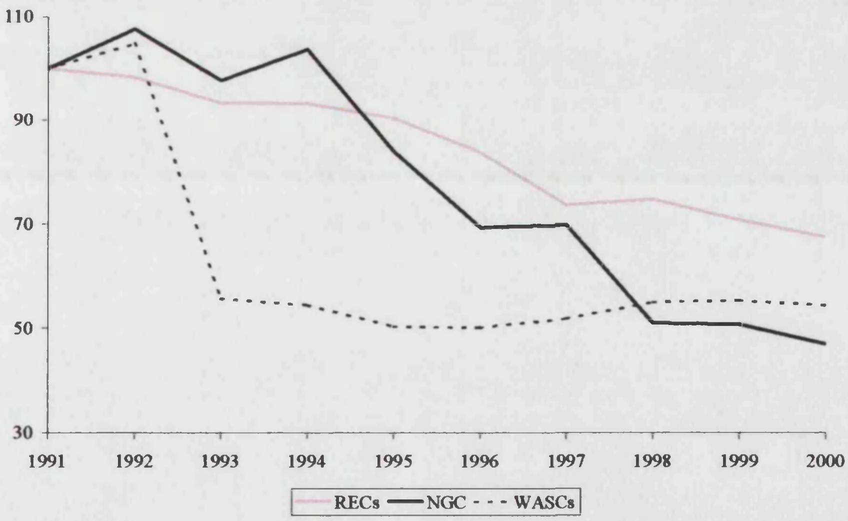

Figure C.7: Non-capital Costs Index

[11]Tilley and Weyman Jones (1999) also use real operating expenditure as an input in the production process. They define this variable as 'revenue minus operating profit, current cost depreciation, and exceptional items'.

Note on labour level

Productivity is usually measured using capital and labour as the main inputs in the production process. The labour level might be measured by the number of employees or by the level of employment costs. There are several problems with using the employment level for these regulated firms however.

- Data on employment costs and number of employees are not available for the regulated distribution companies. Any data used would therefore only be a proxy for the required information. The most frequently used proxy is the data available for the Group business which undertakes several other activities. Trends in this data do not reflect changes in the labour input for distribution. For example, if a company sold its supply business there would be a drop in the labour level of the Group without there being any change in the distribution labour level. This information therefore has the potential to be misleading, particularly given the amount of restructuring which many of the Groups have been though since privatisation.

- Data on employment costs and number of employees is available for NGC and the water companies. The cost data is published in the regulated accounts, and information on the number of employees is available from the regulator (water) or from the Group accounts (NGC). There is a significant concern, cited by several authors, that this data does not accurately reflect the labour input into the production process. This is because a large proportion of activities undertaken by these businesses are now contracted out. A reduction in direct labour hired may therefore be partially offset by an increase in the cost of hiring a contracted worker.

- Other non-capital factors of production (e.g. materials), which are not published separately by the electricity firms, would be omitted from the analysis.

For these reasons we argue that controllable operating costs is a more suitable measure of the non-capital inputs used in the production process. We recognise that the nature by which this variable is calculated, particularly in the electricity sector, means that some of these costs may not reflect actual inputs in the production process. It is the best available proxy, however, given the data provided in the regulatory accounts.

Appendix D

Production function estimation

In this appendix we discuss the different econometric techniques which are used to estimate the parameters of the Cobb-Douglas production function in section 4.4.1. The merits and problems of each approach are outlined.

We estimate production functions for the electricity distribution and water sectors. We do not estimate a separate production function for NGC (electricity transmission) because of the small sample size[1]. We assume that each firm within a sector has the same technology. The coefficients of the estimated *sectoral* production functions are then used to calculate firm-specific productivity growth rates in section 4.5.

We estimate a standard production function and a constant returns to scale (CRS) production function for each sector. The specific equations which we estimate are:

$$y_{it} = a + \alpha k_{it} + \beta l_{it} + \eta_i + \varepsilon_{it} \qquad \text{Standard model}$$

$$\begin{aligned} y_{it} &= a + \alpha k_{it} + (1-\alpha) l_{it} + \eta_i + \varepsilon_{it} \qquad \text{CRS model} \\ &\Rightarrow (y_{it} - l_{it}) = a + \alpha \left(k_{it} - l_{it}\right) + \eta_i + \varepsilon_{it} \end{aligned}$$

where y is output, k is capital, l is controllable operating expenditure, η is a firm-specific effect, and ε is a random error term. All variables are in logs format.

The functions are estimated using a panel data series for each sector. In the electricity distribution sector we have data on twelve regional electricity companies ($N^R = 12$) over a ten year period (1990-2000) ($T^R = 10$). In the water sector we have data on ten water and sewerage companies ($N^W = 10$) over the same ten year period ($T^W = 10$).

The statistical properties of the panel data estimators are based on the assumption that the cross section dimension (N) is large relative to the time series dimension (T).

[1] Only ten years of data is available for one company.

Wooldridge (2002) argues that 'if N is sufficiently large relative to T, and we can assume rough independence in the cross section, then our asymptotic analysis should provide suitable approximation' for finite sample properties. If, however, 'T is of the same order as N...an asymptotic analysis that make explicit assumptions about the nature of the time dependence is needed'. In our datasets we have small and finite N and T. In addition, N and T are close to each other in both sectors; indeed the same in water. We therefore cannot be sure that the properties of the estimators hold here. We recognise this at the outset and note that the estimated coefficients may be inconsistent and/or inefficient with our datasets.

The estimating techniques used are:

- Ordinary least squares on the levels model (OLS);
- Ordinary least squares on the first-differenced model (FDOLS);
- Within Group estimation (WG); and
- Anderson-Hsiao estimation (AH).

In each case, the production function is estimated with time dummies. These capture aggregate factors that affect output over time in the same way for all firms. The year dummies are their own instruments when instruments are used in the estimation procedure.

We are seeking the estimation technique which provides the most reasonable estimates for the capital and labour coefficients, given our datasets. By *reasonable* we mean that:

- the estimates are, according to econometric theory, consistent and efficient;
- the coefficient estimates are consistent with economic theory - ie, they lie between zero and one, and the values are neither very low or very high;
- the estimates are individually and jointly statistically significant - in all tables a * indicates that the estimated coefficient is significant. An estimate is considered significant here if the p-value associated with the coefficient is ≤ 0.01. Standard errors are provided in parentheses; and
- the estimated model yields a high goodness-of-fit measure.

D.1 Data analysis

Before discussing the estimation techniques we take a closer look at the variables in the Cobb-Douglas production function.

D.1.1 Multicollinearity

We find that there is relatively low positive correlation, 22.6%, between the capital and labour variables in the electricity distribution sector. It is significantly higher, 71.25%, for the water sector. This suggests that multicollinearity will not be a major concern in the electricity distribution model, but it may be a concern in the water sector model. In this case the coefficient estimates will be unreliable as they do not measure the direct relationship between the individual explanatory variable and the dependent variable but also capture the collinearity across variables to some extent. With only one regressor, multicollinearity is not an issue in the constant returns to scale model.

D.1.2 Stationarity

The standard production function

To determine whether the explanatory variables in the production function are stationary we estimate the following equations:

$$\omega_{it} = \lambda^{\omega}\omega_{it-1} + v_{it}^{\omega}$$

where $\omega = \{y, k, l\}$.

The results from the OLS regressions are presented in Table D.1. We find that there is a high degree of correlation between each variable and its lagged value, suggesting that the variables are non-stationary. That is, the variables are persistent and a shock in one period will continue to have an impact because of the relationship between current and future values of these variables.

Table D.1: Correlation of variables over time

	Electricity	Water
Output	1.001*	1.001*
	(.0002)	(.0002)
Capital	1.001*	1.001*
	(.0005)	(.0005)
Labour	.9895*	.9822*
	(.0035)	(.0056)

The production function with non-stationary regressors becomes:

$$\begin{aligned} y_{it} &= a + \alpha\left(\lambda^k k_{it-1} + v_{it}^k\right) + \beta\left(\lambda^l l_{it-1} + v_{it}^l\right) + \eta_i + \varepsilon_{it} \\ &= a + \alpha\lambda^k k_{it-1} + \beta\lambda^l l_{it-1} + \eta_i + \left(\varepsilon_{it} + \alpha v_{it}^k + \beta v_{it}^l\right) \end{aligned}$$

This function is dynamic in nature. The capital and labour variables in period t are correlated with the error term in period t-1. They are also potentially correlated with a firm-specific fixed effect. We therefore have a problem with endogenous regressors in our model.

The constant returns to scale production function

We also consider the variables in the model with constant returns to scale technology. We estimate the following dynamic model for each variable:

$$\omega_{it} = \lambda^{\omega}\omega_{it-1} + v_{it}^{\omega}$$

where $\omega = \{(y - l), (k - l)\}$.

The results from the OLS regressions are presented in Table D.2. We again find that there is a high degree of correlation between each variable and its lagged value. That is, the variables are non-stationary. This means that the constant returns to scale production function should also be modelled as a dynamic function with an endogenous regressor.

Table D.2: Correlation of CRS variables over time

	Electricity	Water
Output/Labour	1.011*	1.017*
	(.0033)	(.0067)
Capital/Labour	1.016*	1.014*
	(.0053)	(.0053)

Summary

We have shown that, in our production function, the regressors are non-stationary (persistent). These variables are therefore correlated with the error term, and we need to use estimation techniques which take account of this endogeneity. Where there is a firm-specific effect in the model, it may also be correlated with the capital and labour coefficient over time. These conclusions are true for both the standard production function and the constant returns to scale production function.

D.2 Ordinary least squares

We first estimate the production function using the standard pooled Ordinary Least Squares (OLS) estimation technique. Our pooled OLS estimators allow for the fact that observations are independent across firms, but are not necessarily independent within a firm over time[2]. This estimation procedure provides robust standard error estimates.

We summarise here the main theoretical properties of the pooled OLS estimates.

1. OLS is the Best Linear Unbiased Estimator (BLUE) if:

   (a) $\varepsilon_{it} \sim (0, \sigma^2)$ - ie, there is no autocorrelation and the error term is homoscedastic.

   (b) The regressors, k_{it} and l_{it}, are strictly exogenous (ie, not correlated with ε_{it} $\forall t$).

   (c) The regressors are linearly independent (ie, no multicollinearity) and, hence, there is a unique least square estimator.

   These conditions are referred to as the assumptions of the Classical Linear Regression Model.

2. The OLS estimates are consistent if:

   (a) the regressors and the error term are contemporaneously uncorrelated[3]; and

   (b) there is perfect linear dependencies among the regressors.

3. If the error term exhibits homoscedasticity and if there is no autocorrelation the usual test statistics are valid.

4. If there is evidence of heteroscedasticity in the error term, the estimates are unbiased and consistent but not efficient. We get biased variance estimators and hence standard test statistics are invalid. The problem can be overcome by using generalized least squares (GLS) or by adjusting the standard errors for the presence of heteroscedasticity. According to Wooldridge (2002) 'Lately, it has become more popular to estimate β by OLS even when heteroskedasticity is suspected but to adjust the standard errors and test statistics so that they are valid in the presence of arbitrary heteroskedasticity'. In line with this, our standard error estimates are adjusted so that they are robust to the presence of heteroscedasticity.

5. If there is evidence of autocorrelation in the error term, the estimates are unbiased and consistent but inefficient for all sample sizes. This gives biased goodness-of-fit and test statistics.

6. If the variables in the regression are non-stationary, the OLS estimates have sampling distributions with unknown properties, and the regression coefficients

[2] The ©Stata command used is: xi: reg y k l i.yr, cluster(firm)

[3] Regressors in previous periods can be correlated with the error term in period t.

tend to appear spuriously significant. The OLS estimator is inconsistent, the distribution of R^2 is nondegenerate, the t-distribution diverges and, hence, there is no asymptotically accurate t-values under standard theory (see Wooldridge (2002)).

7. Pooled OLS ignores the firm-specific fixed effect. If the regressors are exogenous and uncorrelated with the firm-effect (ie, the random effects model holds), the OLS estimates are unbiased and consistent but they are inefficient. The standard errors are also biased. If the regressors are correlated with the firm-effect (ie, the fixed effect model holds), the OLS estimates are biased and inconsistent.

8. If there is correlation between the error term and the regressors, the parameters are not identified using OLS. In particular, if the regressors are endogenous the OLS estimate is biased and it is not consistent. We are also unable to carry out standard inference as we no longer know that the estimates are normally distributed.

D.2.1 Estimated coefficients

We carry out the pooled OLS estimation for the standard production function and for the production function with constant returns to scale. Time dummies are included. The results are presented in Table D.3.

In the electricity distribution sector the OLS coefficient on labour is quite low and it is also statistically insignificant. The constant returns to scale model provides more reasonable results. The OLS coefficients on the water sector production function are satisfactory, although we have an unlikely scenario where the labour coefficient is higher than the capital coefficient. In the constant returns to scale model the coefficients are statistically significant but the coefficient value is quite low.

Table D.3: OLS Estimates

	Electricity	Water
Standard model		
Constant	3.255*	2.601*
	(.7494)	(.4991)
Capital coefficient	.7256*	.3294*
	(.1270)	(.0605)
Labour coefficient	.1859	.5177*
	(.1053)	(.1537)
R-squared	0.81	0.91
F-statistic	n.a.	n.a.
Observations	120	100
CRS model		
Constant	2.748*	1.952*
	(.3176)	(.2252)
Capital/Labour coefficient	.7515*	.3087*
	(.1101)	(.0607)
R-squared	0.86	0.72
F-statistic	130.31	n.a.
Observations	120	100

D.2.2 Testing for heteroscedasticity

In both sectors the firms are of different sizes and hence we suspect that there may be heteroscedasticity in the model. Figures D.1 to D.4 plot the squared OLS residuals (X-axis) against the fitted value of output (Y-axis) for each of the estimated models. We see no evidence of a systematic relationship between these variables. This suggests that the error terms are homoscedastic.

We also carry out the White Test on the null hypothesis that the error terms are homoscedastic[4]. The results are presented in Table D.4. We reject the null hypothesis for the standard production functions at the 1% significance level and therefore find that the error terms may be heteroscedastic. Our robust standard errors take account of the potential for heteroscedasticity in the model and provide a reliable basis for making inferences. In the constant returns to scale model, the null hypothesis of homoscedasticity is accepted.

[4]In the White Test the squared residuals are regressed on all explanatory variables, their cross-products and their squared values. The null hypothesis of no heteroscedasticity is rejected if nR^2 from this regression is greater than the chi-squared critical value. This is because the error term is heteroscedastic if there is a relationship - i.e. a high R^2 - between the residuals and the explanatory variables.

Table D.4: OLS: White Test Results

	Electricity		Water	
	Statistic	H_0	Statistic	H_0
Standard model	37.18	Reject	70.52	Reject
CRS model	27.45	Accept	26.92	Accept

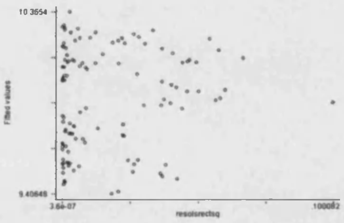

Figure D.1: Electricity and OLS: relationship between the squared residuals and the fitted values

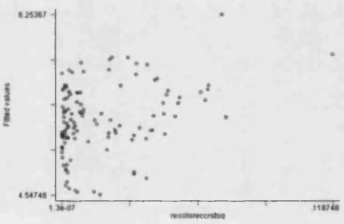

Figure D.2: Electricity and OLS: relationship between squared residuals and fitted values in the CRS model

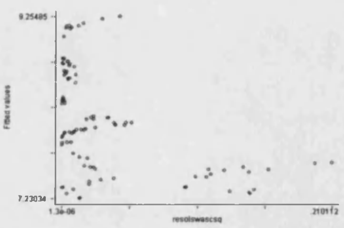

Figure D.3: Water and OLS: relationship between squared residuals and fitted values

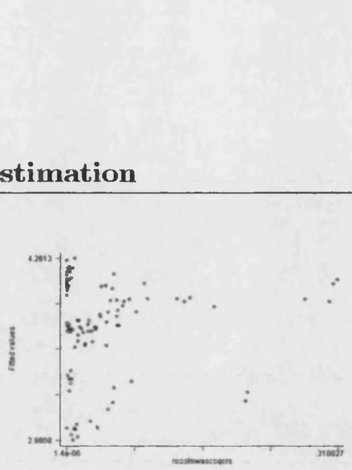

Figure D.4: Water and OLS: relationship between squared residuals and fitted values in OLS model

Testing for autocorrelation

We expect to get serial correlation in the errors as we have an omitted time-constant variable - the firm-specific effect - in the production function. Figures D.5 to D.8 plot the OLS residual (Y-axis) against its lagged value (X-axis) for each of the models. There is concern that the error term in the regression is autocorrelated in both the electricity distribution and water sectors.

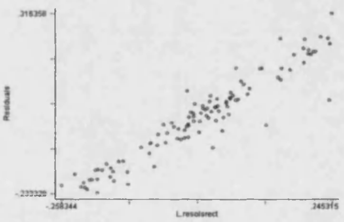

Figure D.5: Electricity and OLS: relationship between residuals and lagged residuals

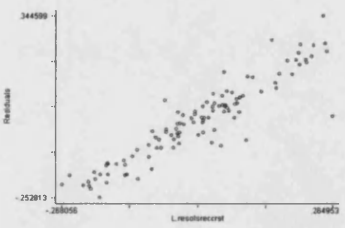

Figure D.6: Electricity and OLS: relationship between residuals and lagged residuals in the CRS model

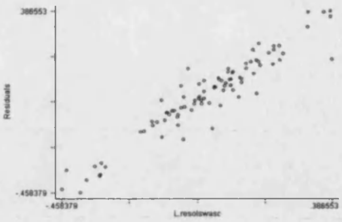

Figure D.7: Water and OLS: relationship between residuals and lagged residuals

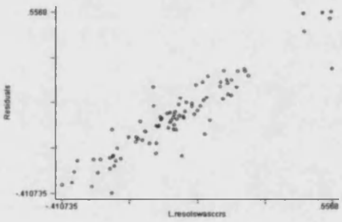

Figure D.8: Water and OLS: relationship between residuals and lagged residuals in CRS model

We determine the degree of autocorrelation by estimating the following equation for each sector.

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + v_{it}$$

The results are presented in Table D.5 and confirm a high degree of correlation in the residuals over time. The degree of autocorrelation, although still very significant, is lowest in the model with constant returns to scale.

Table D.5: Autocorrelation in the OLS models

$\widehat{\rho}$	Electricity	Water
Standard model	.9604*	.9341*
	(.0305)	(.0331)
CRS model	.9364*	.9180*
	(.0369)	(.0383)

D.2.3 Analysis

Our main conclusions on the OLS estimates of our production functions are as follows.

1. The constant returns to scale model provides reasonable estimates for the coefficients of the production functions in both sectors. The estimated coefficients in the standard production function are less useful.

2. The estimated residuals appear to be autocorrelated. This means that the OLS estimators are unbiased but not efficient. Inferences based on the standard errors are unreliable.

3. We have non-stationary regressors and an autocorrelated error term. The regressors are therefore correlated with the error term over time. In this situation, the OLS estimates are biased upwards (see Nickell (1981)) and inconsistent.

4. The firm-effect has not been captured by this estimation technique. It is essentially treated as a missing variable in the error term. If it is correlated with the regressors, the pooled OLS estimates are biased and inconsistent. If the fixed effect is not correlated with the regressors, the estimates are unbiased and consistent but they are inefficient relative to generalized least squares (GLS).

D.3 OLS on the first-differenced model

In this section we first-difference the production function and remove the firm-specific effect.

$$
\begin{aligned}
(y_{it} - y_{it-1}) &= \alpha (k_{it} - k_{it-1}) + \beta (l_{it} - l_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1}) \quad \text{FD Model} \\
&\Rightarrow \Delta y_{it} = \alpha \Delta k_{it} + \beta \Delta l_{it} + \Delta \varepsilon_{it}
\end{aligned}
$$

The first-differenced constant returns to scale production function is:

$$
\Delta (y_{it} - l_{it}) = \alpha \Delta (k_{it} - l_{it}) + \Delta \varepsilon_{it} \qquad \text{CRS FD Model}
$$

The first-differenced production functions are estimated using pooled OLS and time dummies are included in the models. Our estimates allow for the fact that observations are independent across firms, but are not necessarily independent within a firm over time[5]. This procedure provides robust estimators of variance.

We first summarise the main theoretical properties of the OLS estimates of the coefficients in the first-differenced production function. The properties are the same as those discussed in section D.2, except now the discussion relates to the first-differenced variables rather than their levels counterparts.

1. The estimates are unbiased and consistent if the first-differenced regressors are uncorrelated with the first-differenced error terms (weak exogeneity), and if the first-differenced regressors are linearly independent (ie, no multicollinearity).

[5]The Stata command used is: xi: reg Δy Δk Δl i.yr, cluster(firm)

2. If there is correlation (contemporaneous or over time) between the first-differenced error term and the first-differenced regressors, the estimates are inconsistent.

3. If the error term in the levels model is autocorrelated, first-differencing reduces the degree of autocorrelation. If the residual has serial correlation, but the first-differenced residuals are serially uncorrelated, these estimates are, according to Wooldridge (2002), the 'most efficient in the class of estimators using the strict exogeneity assumption'. The estimates are less efficient than within group estimates when the residuals are serially correlated.

4. Differencing eliminates the firm-specific effect. If the firm-effect is correlated with the regressors, the estimates are unbiased and consistent. They are less efficient than the within group estimates however. If the firm-effect is independent of the regressors, the estimates are unbiased and consistent, but they are inefficient relative to generalised least squares (GLS).

5. If there is not much variation in the data over time, the first-difference will be identically zero, and the regressors will not be linearly independent. The estimated coefficients will therefore be 'imprecise' (Wooldridge, 2002) if the regressors do not vary much over time.

6. First-differencing induces stationarity in the model which can reduce the probability of spurious results.

D.3.1 Estimated coefficients

We estimate the first-differenced standard production function and the first-differenced constant returns to scale production function. In both cases time dummies are included in the regression. The results are presented in Table D.6.

The estimated coefficients are not very useful. We get negative labour coefficients in both sectors for the standard production function. These clearly conflict with economic theory. The constant returns to scale production functions have extremely high capital coefficients, implying a negligible coefficient on labour. These values are considered unrealistic.

Table D.6: FDOLS Estimates

	Electricity	**Water**
Standard model		
Capital coefficient	.0255	-.0646
	(.0298)	(.0693)
Labour coefficient	-.0139	-.0047
	(.0079)	(.0062)
R-squared	0.60	0.44
F-statistic	727.34	n.a.
Observations	108	90
CRS model		
Capital/Labour coefficient	.9958*	.9818*
	(.0115)	(.0099)
R-squared	0.98	0.99
F-statistic	3615.04	n.a.
Observations	108	90

D.3.2 Testing for heteroscedasticity

Figures D.9 to D.12 plot the squared OLS residuals (X-axis) against the fitted value of output (Y-axis) for each of our models. In all cases there seems to be a relationship between the squared residual and the fitted value, suggesting that the variance of the residual is not constant. Heteroscedasticity may therefore be present in this model.

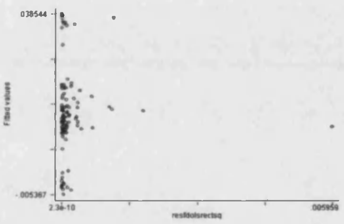

Figure D.9: Electricity and FDOLS: relationship between squared residuals and fitted values

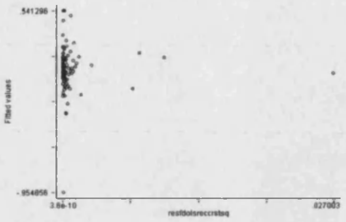

Figure D.10: Electricity and FDOLS: relationship between squared residuals and fitted values in the CRS model

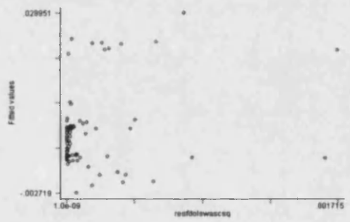

Figure D.11: Water and FDOLS: relationship between squared residuals and fitted values

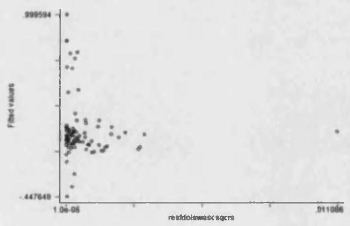

Figure D.12: Water and FDOLS: relationship between squared residuals and fitted values in CRS model

We also carry out the White Test on the null hypothesis that the error terms are homoscedastic. The results are presented in Table D.7. We accept the null hypothesis of homoscedasticity for the electricity distribution and water production functions.

Table D.7: FDOLS White Test Results

	Electricity		Water	
	Statistic	H_0	Statistic	H_0
Standard model	25.31	Accept	25.19	Accept
CRS model	10.03	Accept	16.50	Accept

D.3.3 Testing for autocorrelation

Autocorrelation appears to be less of a problem in the first-differenced model than in the levels model. Figures D.13 to D.16 plot the residual (Y-axis) against its lagged value (X-axis) for each of our models. There is no distinct relationship between these variables in any of the models.

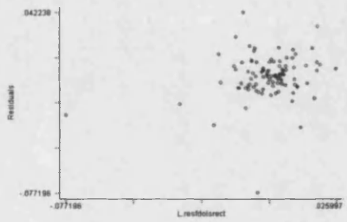

Figure D.13: Electricity and FDOLS: relationship between residuals and lagged residuals

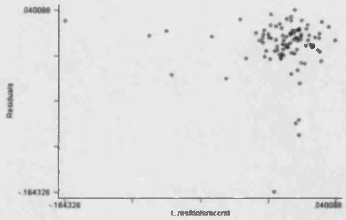

Figure D.14: Electricity and FDOLS: relationship between residuals and lagged residuals in the CRS model

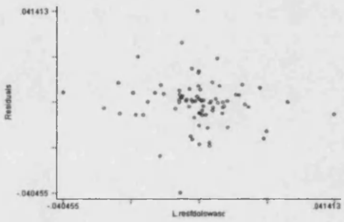

Figure D.15: Water and FDOLS: relationship between residuals and lagged residuals

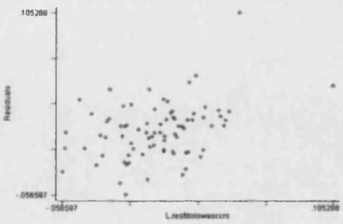

Figure D.16: Water and FDOLS: relationship between residuals and lagged residuals in CRS model

We determine the degree of autocorrelation by estimating the following equation for each sector.

$$\Delta\varepsilon_{it} = \rho\Delta\varepsilon_{it-1} + v_{it}$$

The results are presented in Table D.8. We find that there is not a significant degree of correlation between the error term and its lagged value in any of the first-differenced models, with the highest correlation in the water constant returns to scale model. In most cases the estimate of ρ is not statistically significant.

Table D.8: FDOLS: autocorrelation

$\widehat{\rho}$	Electricity	Water
Standard model	.2498	-.0951
	(.1147)	(.104)
CRS model	-.0448	.3339*
	(.1037)	(.0954)

D.3.4 Analysis

Our main conclusions on the OLS estimates for the first-differenced production functions are as follows.

1. None of the production function models provide reasonable coefficient estimates.

2. The fixed effect is removed when we first-difference the model. The estimates are therefore unbiased and consistent if the assumptions of the Classical Linear Regression Model apply to the first-differenced variables.

3. Heteroscedasticity does not appear to be a significant issue for the electricity distribution or water production functions.

4. There is some evidence of autocorrelation in our models, but it is not as significant as in the levels model. There is thus some loss of efficiency.

5. The first-differenced variables in the constant returns to scale model are persistent[6]. Given the absence of any significant autocorrelation in the error term, we conclude that the regressor may be endogenous. With endogenous regressors in the model the estimates are biased and inconsistent.

This summary suggests that the OLS estimates of the first-differenced models should provide us with more reliable estimates than the OLS estimates of the levels

[6]The first-differenced variables are linear combination of persistent variables.

model. We find, however, that with our datasets the estimates of the first-differenced model are, in the main, unusable. This may suggest that problems are arising from the small sample size, as the standard predictions of econometric theory are largely based on asymptotic assumptions. It may also reflect the fact that the regressors are endogenous in these models.

D.4 Within Group estimation

In this section we treat the firm-effect as an unknown parameter which is to be estimated. No assumptions are made about its distribution, or about its relationship with the regressors. In particular, the parameter may be non-random, and it may be correlated with capital and/or labour. This is referred to as the fixed effects model.

The assumptions of the fixed effects model are:

- The error term has zero mean: $E(\varepsilon_{it}) = 0 \; \forall i, t$
- The error term is not correlated over time: $E(\varepsilon_{it}, \varepsilon_{is}) = 0 \; \forall t \neq s, i$
- The regressors are strictly exogenous (uncorrelated with the error term across time and contemporaneously): $E(k_{it}, \varepsilon_{is}) = E(l_{it}, \varepsilon_{is}) = 0 \; \forall t, s, i$
- The error term is homoscedastic across individuals and time: $E(\varepsilon_{it}) = \sigma_{\varepsilon}^2$

Inferences are made conditional on the effects which arise in the sample.

We estimate the parameters of the fixed effects models using the within group estimation technique (WG). The fixed effect is eliminated by taking deviations from time means and carrying out pooled OLS on the following equation[7]:

$$(y_{it} - \overline{y_i}) = \alpha(k_{it} - \overline{k}_i) + \beta(l_{it} - \overline{l}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

where $\overline{y_i} = \frac{\sum_t y_{it}}{T}$; $\overline{k}_i = \frac{\sum_t k_{it}}{T}$; $\overline{l}_i = \frac{\sum_t l_{it}}{T}$; and $\overline{\varepsilon}_i = \frac{\sum_t \varepsilon_{it}}{T}$.

This regression yields the coefficient estimates $\widehat{\alpha}_w$ and $\widehat{\beta}_w$. We can estimate the constant term, a, by carrying out the following regression:

$$\overline{y} = a + \widehat{\alpha}_w \overline{k} + \widehat{\beta}_w \overline{l} + \overline{\varepsilon}$$

where $\overline{y} = \frac{\sum_i \sum_t y_{it}}{TN}$; $\overline{k} = \frac{\sum_i \sum_t k_{it}}{TN}$; $\overline{l} = \frac{\sum_i \sum_t l_{it}}{TN}$; and $\overline{\varepsilon} = \frac{\sum_i \sum_t \varepsilon_{it}}{TN}$. This regression gives us an estimate $\widehat{a}_w$. The fixed-effect can then be estimated using the averages over time. That is:

[7]The ©Stata command used is xi: xtreg y k l i.yr, fe

$$\begin{aligned} \overline{y_i} &= a + \alpha\overline{k}_i + \beta\overline{l}_i + \eta_i + t + \overline{\varepsilon}_i \\ &\Rightarrow \widehat{\eta}_i = \overline{y_i} - \widehat{a}_w - \widehat{\alpha}_w\overline{k}_i - \widehat{\beta}_w\overline{l}_i \end{aligned}$$

We summarise the main theoretical properties of the within group estimates here.

1. If the firm-specific effect is non-random (ie, the fixed effects model is valid), the estimator is BLUE, assuming all other elements of the classical linear regression model hold. The assumptions need to hold for the residual term, $(\varepsilon_{it} - \overline{\varepsilon}_i)$, and the regressors, $(k_{it} - \overline{k}_i)$ and $(l_{it} - \overline{l}_i)$. Pooled OLS and GLS yield biased and inconsistent estimates in this situation.

2. If the firm-specific effect is uncorrelated with the regressors (ie, the random effects model holds), the within group estimator is consistent but it is less efficient than GLS. If the variation in the variables is across firms, rather than over time for a given firm, the loss of efficiency can be large.

3. The within group estimator is efficient if the error term has a constant variance across t and if it is serially uncorrelated. The efficiency is reduced if either of these conditions don't hold. In addition, the estimated standard errors will be misleading, and inference based on standard tests will not be valid. Consistency of the estimate is not affected.

4. The estimator is consistent when the regressors are strictly exogenous (conditional on η_i), and N, T or NT tend to infinity. If N is fixed and T tends to infinity, the estimator for the coefficients on the regressors are consistent, but the estimator for the constant is not consistent.

5. If the regressors are not strictly exogenous the estimator is inconsistent. The size of the bias reduces as T increases. According to Wooldridge (2002) the bias 'can be sizeable' if the degree of correlation in the error term over time is close to 1 and/or if 'the process $\{x_{it}\}$ has very persistent elements' (where x_{it} is a regressor). The bias will be lower than with pooled OLS on the first-differenced equation as $N \to \infty$.

D.4.1 Estimated coefficients

We carry out the within group analysis for the standard production function and for the production function with constant returns to scale. In both cases we estimate the equations with time dummies. The estimated coefficients are presented in Table D.9.

The standard production function estimates are not reasonable for the electricity distribution or water sectors. In the electricity distribution case we get a negative coefficient on labour. In the water sector we get a negative coefficient on capital. The water constant returns to scale coefficient is also unreasonable as it is greater than one. The coefficient estimate in the electricity constant returns scale model is more reasonable, although it is very high. This outcome is similar to that observed when we estimated the production functions using pooled OLS on the first-differenced model. We note that the F-statistic is higher for the constant returns to scale model in both sectors, suggesting that this restriction is appropriate.

Table D.9: Within Group Estimates

	Electricity	**Water**
Standard model		
Capital coefficient	.1154	-.0926
	(.0553)	(.0384)
Labour coefficient	-.0134	.0045
	(.0141)	(.0116)
Constant	9.006*	8.876*
	(.4405)	(.3268)
R-squared (overall)	0.45	0.69
F-statistic	42.49	39.03
Observations	120	100
CRS model		
Capital/Labour coefficient	.9558*	1.040*
	(.0255)	(.0415)
Constant	2.135*	-1.123*
	(.0775)	(.1751)
R-squared (overall)	0.85	0.51
F-statistic	406.75	461.99
Observations	120	100

D.4.2 Testing for heteroscedasticity

Figures D.17 to D.20 plot the squared within group residuals (X-axis) against the fitted value of output (Y-axis) for each of our models. With this estimation technique the fitted value is:

$$\widehat{y}_{it} = \widehat{a} + \widehat{\alpha}k_{it} + \widehat{\beta}l_{it} + \widehat{\eta}_i$$

and the residual is:

$$\widehat{\varepsilon}_{it} = y_{it} - \widehat{y}_{it}$$

We see that there is, in most cases, a systematic relationship between the fitted value and the squared residual, with low values of the squared residual for all fitted values. This suggests that there may be heteroscedasticity in our models, although the relationship is less clear in the water sector than in the electricity distribution sector.

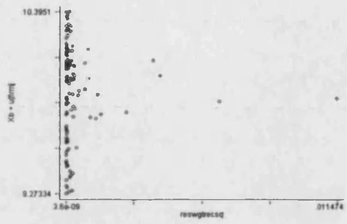

Figure D.17: Electricity and WG: relationship between squared residuals and fitted values

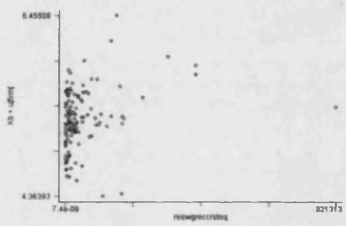

Figure D.18: Electricity and WG: relationship between squared residuals and fitted values in the CRS model

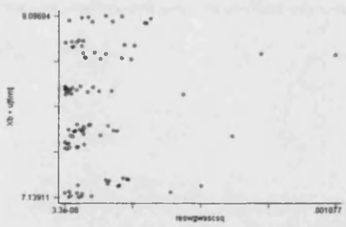

Figure D.19: Water and WG: relationship between squared residuals and fitted values

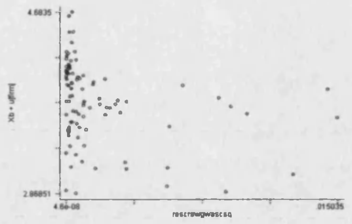

Figure D.20: Water and WG: relationship between squared residuals and fitted values in CRS model

D.4.3 Testing for autocorrelation

In all sectors there is concern that the error term in the regression is autocorrelated, although the concern is least obvious for the standard water production function. This is demonstrated in Figures D.21 to D.24 which plot the residual (Y-axis) against its lagged value (X-axis) for each of our models.

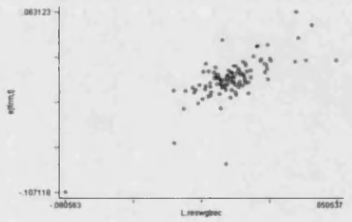

Figure D.21: Electricity and WG: relationship between residuals and lagged residuals

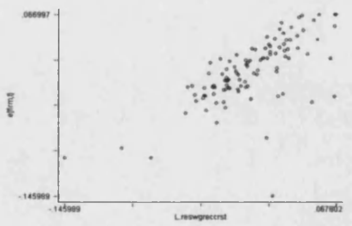

Figure D.22: Electricity and WG: relationship between residuals and lagged residuals in the CRS model

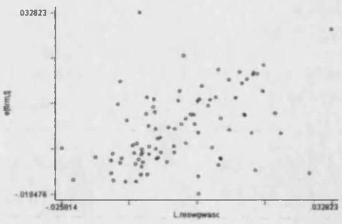

Figure D.23: Water and WG: relationship between residuals and lagged residuals

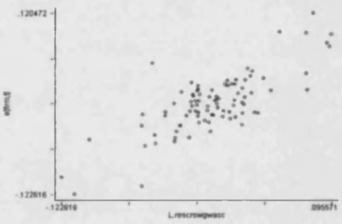

Figure D.24: Water and WG: relationship between residuals and lagged residuals in the CRS model

We also test for autocorrelation by running the following regression for each sector:

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + v_{it}$$

The results are presented in Table D.10. We find that there is evidence of autocorrelation in all the models. The lowest level of autocorrelation is in the water standard production function.

Table D.10: Within group: autocorrelation

$\widehat{\rho}$	Electricity	Water
Standard model	.9718*	.4175*
	(.0885)	(.0866)
CRS model	.7452*	.7820*
	(.0767)	(.0678)

D.4.4 Analysis

Our main conclusions on the within group estimates of our production functions are as follows.

1. With our data, the within group approach provides unreasonable coefficient estimates for the standard production function and the constant returns to scale production function in both sectors.

2. We expect the fixed effect to be correlated with our non-stationary regressors. The fixed effect model is therefore valid, and the within group estimates are BLUE if assumptions of the Classical Linear Regression model hold.

3. The error term is potentially autocorrelated and heteroscedastic. The within group estimates will therefore be consistent but inefficient, and the standard errors will be biased.

4. The variation in our capital and labour data primarily occurs across firms rather than over time. We therefore expect a loss of efficiency in the within group estimates.

5. The error term is autocorrelated, and the capital and labour variables are generated from a dynamic process. We therefore have endogenous regressors in the model and the within group estimates are biased and inconsistent. With small T, and persistent regressors, the bias may be quite large.

D.5 Anderson-Hsiao estimation

We know that the production function model is dynamic in nature. This is because the regressors are non-stationary, and there is evidence that the error terms are autocorrelated in the fixed-effect model. The regressors in our estimation are therefore endogenous. A similar conclusion was reached by Blundell and Bond (1999) who found that, in many cases, capital and non-capital inputs are non-stochastic and, hence, there is correlation between the error term and the regressors over time.

In this situation the within group estimator has a negative bias but it is consistent. The OLS estimator on the levels model has a positive bias, and the OLS estimator on the first-differenced model has a negative bias. Both OLS estimators are inconsistent. The size of these biases are estimated by Nickell (1981). Given the problems with these estimators, Wooldridge (2002) suggests that the production function with endogenous regressors can be estimated by transforming the equation to remove η_i, and using instrumental variables to estimate the transformed equation.

The standard transformations considered are first-differencing, within group (taking deviations from time means) and GLS (taking weighted deviations from time means). The first-differencing approach is primarily associated with Anderson and Hsiao (1982). Wooldridge (2002) argues that, of all the transformations that can be used, 'first differencing is more attractive'. In line with this, we choose to use the Anderson-Hsiao approach to estimate our production functions.

We take first-differences of the production function and eliminate the firm-specific fixed effect. The functions are estimated with time dummies. We use values of the regressors lagged two or more periods as instruments in the Anderson-Hsiao estimation[8]. We assume that the error term is serially independent and that:

$$\begin{aligned} E(k_{it-s}, \Delta\varepsilon_{it}) &= 0 \text{ for } T = 3, ...T, \ s \geq 2 \\ E(l_{it-s}, \Delta\varepsilon_{it}) &= 0 \text{ for } T = 3, ...T, \ s \geq 2 \\ E((k_{it-s} - l_{it-s}), \Delta\varepsilon_{it}) &= 0 \text{ for } T = 3, ...T, \ s \geq 2 \end{aligned}$$

Hsiao (1986) argues that, given the above assumptions, lagged values of the regressors meet the requirements of valid instruments. That is, they are correlated with the variables to be instrumented, and they are not correlated with the error term (by the assumption of weak exogeneity).

We summarise the main theoretical properties of the Anderson-Hsiao estimator here. Many of the properties are common to standard two-stage least squares, or instrumental variables, estimation techniques.

1. If the instruments are uncorrelated with the first-differenced disturbances, and highly correlated with the first-differenced regressors, the Anderson-Hsiao estimates are biased but consistent.

2. The variance of the estimator is greater than the variance of the within group estimator. That is, there is a loss of efficiency. This difference can be minimised by getting the maximum correlation between the instruments and the regressors.

3. If the endogenous regressors are only weakly related to the instruments, the standard errors may be large and the coefficient estimates will be significantly biased. Staiger and Stock (1997) suggest that 'even with large sample sizes, instruments that have small partial correlation with an endogenous explanatory variable can lead to substantial biases in 2SLS' (from Wooldridge, p104).

4. The efficiency of the estimator increases with the number of instruments. According to Wooldridge (2002) 'asymptotically, we always do better by using as many instruments as are available, at least under homoskedasticity'.

5. If there are too many instruments, and if the instruments are weak, there are potentially large biases in the estimates in finite samples. According to Wooldridge (2002) the 'lesson is that, even with a very large sample size and zero correlation between the instruments and error, we should not use too many overidentifying restrictions'.

[8] ©Stata Command: xi: xtivreg y (k,l=instruments) i.yr, fd

6. The estimator is no longer consistent if the errors are serially correlated.

7. Heteroscedasticity does not affect the consistency of the estimator and is only a 'minor nuisance for inference' (Wooldridge, 2002). If heteroscedasticity is present then more efficient estimation is possible. The standard Sargan Test is not valid if there is heteroscedasticity in the model.

8. The Anderson-Hsiao estimates will be inconsistent if the instrument is actually correlated with the error term. The size of the inconsistency gets larger if the instrument is only weakly correlated with the endogenous regressor. According to Wooldridge (2002, p102) 'In such cases is may be better to just use OLS' as it may be 'less inconsistent'.

D.5.1 Choosing instruments

We use values of capital and labour lagged two, three and four periods (various permutations) as instruments when estimating the standard production function. In the constant returns to scale model, we use various permutations of the $(k - l)$ variable lagged two, three and four periods as instruments. We consider here how one set of instruments is chosen over another.

Levels or first-differences

We calculate the Anderson-Hsiao coefficients using lagged values of capital and labour levels. We could also have used lagged first-differences of these variables but we argue that there is no loss of efficiency in using levels rather than first-differences. This is supported by others in the field. Arellano (1989) recommends using levels because when differences are used instead the estimator can have a very large variance. Wooldridge (2002) also notes that there is no loss of efficiency using levels rather than first-differences as 'the latter is a linear combination of the former'.

Number of instruments

The efficiency of the Anderson-Hsiao estimator increases with the number of valid instruments included in the instrument set. Estimation becomes difficult, however, if the instrument set is very large. In addition, as noted by Wooldridge (2002), 'many overidentifying restrictions are known to have poor finite sample properties'. We therefore try to use more than one instrument for each endogenous variable but we limit the number of lags used. This allows us to trade-off the efficiency gains of more instruments against the possibility of increased bias in the estimate. This is consistent with Wooldridge's (2002) suggestion that 'In practice, it may be better to use a couple of lags rather than lags back to t=1'.

Possibility of weak instruments

The most recent lagged values of capital and labour are likely to provide weak instruments for the current values of the first-differenced regressors. Blundell and Bond (1999) discuss the problem of weak instruments in the context of Generalised Methods of Moments (GMM) estimation of the production function. They consider an AR(1) model:

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}$$

The lagged value of y is a weak instrument if 'the value of the autoregressive parameter α increase towards unity; and second, as the variance of the permanent effects (η_i) increases relative to the variance of the transitory shocks (v_{it})'. Under these conditions the 'first-differenced GMM estimator has been found to have poor finite sample properties (bias and imprecision)'. The same argument applies to the Anderson-Hsiao estimator.

Our capital and labour variables are non-stationary, as shown in section D.1. Specifically our estimates for λ^k and λ^l are close to 1 in the equations:

$$\begin{aligned} k_{it} &= \lambda^k k_{it-1} + v_{it}^k \\ l_{it} &= \lambda^l l_{it-1} + v_{it}^l \end{aligned}$$

This implies that the lagged level of each variable has a low correlation with the current first-differenced value:

$$\begin{aligned} (k_{it} - k_{it-1}) &= \left(\lambda^k - 1\right) k_{it-1} + v_{it}^k \\ (l_{it} - l_{it-1}) &= \left(\lambda^l - 1\right) l_{it-1} + v_{it}^l \end{aligned}$$

The recently lagged levels provide weak instruments in the first-differenced standard production function. We therefore prefer to use instrument sets which use a number of lagged levels going back in time. In the constant returns to scale model, the recently lagged levels of $(k - l)$ are also likely to provide weak instruments for $\Delta(k - l)$. We therefore again argue that instrument sets which use lagged levels further back in time will yield more satisfactory estimates.

Sargan Test for instrument validity

The above discussion suggests that we want to use a number of lagged values of capital and labour as instruments for the current first-differenced regressors. We limit our attention to four period lags to avoid biasing the results significantly by using too many instruments.

When we have more instruments than endogenous regressors we can test whether the additional instruments are valid (ie, uncorrelated with the residual) using the Sargan Test. We let Z be the full set of instruments and Z' be a subset of these instruments (one option is $Z' = Z$). Following Wooldridge (2002) we carry out the Anderson-Hsiao regression with the instrument set Z' (which has more instruments than endogenous regressors) and calculate the residual, $\widehat{\varepsilon}_{it}$, from the regression. We then regress this residual onto the full set of instruments, Z, and get the goodness-of-fit measure, R^2. If R^2 is small, the additional instruments are not correlated with the unexplained element of the production function, and we can argue that the instrument set Z' is appropriate.

The Sargan statistic is:

$$S = nR^2 \sim \chi^2_{l-g}$$

The test statistic, S, has a Chi-squared distribution with degrees of freedom equal to the number of variables in the instrument set Z' (l) minus the number of endogenous regressors (g). $l - g$ is the number of overidentifying restrictions.

The null hypothesis being tested is that the instrument set Z' is uncorrelated with the error term, and hence it is a valid instrument set to use. This hypothesis will be rejected if $nR^2 > \chi^2_{g-l}$ for the specified degrees of freedom. The hypothesis will be accepted if $nR^2 < \chi^2_{g-l}$ for the specified degrees of freedom. *The evidence in favour of the null hypothesis is stronger the higher is the p-value on the Sargan statistic.*

If the null hypothesis is accepted when $Z = Z'$ it is better to use the instrument set which uses all of the relevant variables as this increases the efficiency of the estimates. If the null hypothesis is rejected when $Z = Z'$, we need to determine which alternative instrument set - involving fewer variables - should be used. If we have a number of instrument sets, with the same number of instruments, which all satisfy the Sargan test, we choose the set which has the Sargan statistic with the highest p-value. In this case the justification for accepting the instrument set is strongest (ie, the minimum significance level at which the null hypothesis could be rejected is very high).

D.5.2 Coefficient estimation

The electricity distribution sector

The Anderson-Hsiao estimates of the electricity distribution production functions are provided in Table D.11. In the standard model we get coefficient estimates which are inconsistent with production theory. We have negative labour and/or capital coefficients, or capital coefficients which are very high (greater than one in most cases), for all instrument options. This estimation technique therefore provides no reasonable estimates for this data when the standard production function is considered.

In the constant returns to scale model we also find, in most cases, that the estimated coefficients are inconsistent with economic theory. In particular, the estimated coefficient is either very high or greater than 1. We get the most reasonable estimates when three-period lags, and two- and three-period lags, are used as instruments. We find that the Sargan Test is satisfied when two- and three-periods lags are included in the instrument set[9]. A comparison of the goodness-of-fit measures (R^2 and the χ^2 statistic) indicates that it is reasonable to impose the constant returns to scale restriction on the model when the two- and three-period lagged values are used as instruments.

Table D.11: Anderson-Hsiao Estimates for electricity distribution

Standard model $Z = (k, l)$	Lag2	Lag3	Lag4	Lag2+3	Lag2+4	Lag3+4	Lag 2+3+4
α	.2120	7.734	-.0599	.9600	-.1606	-.0844	-.0324
	(1.127)	(69.31)	(1.072)	(2.387)	(.7026)	(.5671)	(.5163)
β	-.0119	.5193	-.1455	.0677	-.0852	-.0198	-.0209
	(.1018)	(4.454)	(.1489)	(.1742)	(.0783)	(.0524)	(.0456)
a	.0155*	.01639	.0092	.0162*	.0142	.0171*	.0163*
	(.0041)	(.0471)	(.0146)	(.0058)	(.0085)	(.0067)	(.0061)
R^2(overall)	0.76	0.78	0.43	0.78	0.76	0.77	0.67
χ^2	19.57	0.11	4.93	4.32	9.05	13.86	14.22
S	n.a.	n.a.	n.a.	0.29	1.82	1.12	1.15
S p-value	n.a.	n.a.	n.a.	0.87	0.40	0.57	0.86
Observations	84	72	60	72	60	60	60
CRS model $Z = (k - l)$	Lag2	Lag3	Lag4	Lag2+3	Lag2+4	Lag3+4	Lag 2+3+4
α	.9635*	.8272*	1.099*	.9211*	1.024*	.9782*	.9734*
	(.0968)	(.2194)	(.1529)	(.0880)	(.0853)	(.0994)	(.0751)
a	.0168*	.0225	-.0039	.0169*	.0016	.0050	.0053
	(.0059)	(0.142)	(.0125)	(.0067)	(.0077)	(.0086)	(.0071)
R^2(overall)	0.82	0.82	0.81	0.82	0.81	0.81	0.81
χ^2	321.81	93.64	156.12	254.57	288.5	240.58	310.34
S	n.a.	n.a.	n.a.	0.15	0.42	0.36	0.35
S p-value	n.a.	n.a.	n.a.	0.70	0.52	0.55	0.84
Observations	84	72	60	72	60	60	60

We conclude that, given our dataset, the preferred model for the electricity distribution sector is the constant returns to scale model. The coefficient on capital is

[9]The p-value on the Sargan statistic is high (0.7). This indicates that there is sufficient evidence in favour of the null hypothesis that the instrument set is valid.

estimated using the two- and three-period lagged values of $(k - l)$ *as instruments.* We explore the properties of the residual in this model to test for autocorrelation and heteroscedasticity.

Figure D.25 plots the squared Anderson-Hsiao residuals (X-axis) against the fitted value of output (Y-axis). There seems to be a relationship between the squared residual and the fitted value, suggesting that the variance of the residual is not constant. Heteroscedasticity may therefore be present in this model. This does not affect the consistency of the estimates but the Sargan Test results may not be valid.

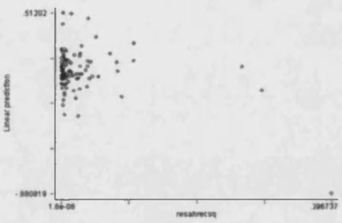

Figure D.25: Electricity and Anderson-Hsiao estimation: relationship between squared residuals and fitted values in the CRS model

Figure D.26 plots the residuals (Y-axis) against their lagged values (X-axis). There is no distinct relationship between these variables. We also run the following regression:

$$\Delta\varepsilon_{it} = \rho\Delta\varepsilon_{it-1} + v_{it}$$

The estimated coefficient, $\widehat{\rho} = -0.001$, indicates that there is no correlation between the error term and its lagged value. We therefore conclude that autocorrelation is not a problem in this model.

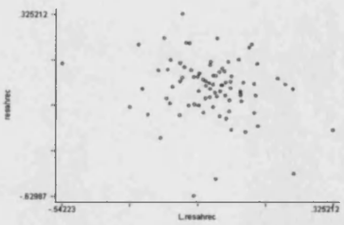

Figure D.26: Electricity and Anderson-Hsiao estimation: relationship between residuals and lagged residuals in the CRS model

The water sector

The Anderson-Hsiao estimates of the water sector production function are provided in Table D.12. With the standard production function we, again, get coefficient estimates which are inconsistent with production theory. We get negative labour and/or capital coefficients for all instrument options. In addition, the coefficient estimates are statistically insignificant. This estimation technique therefore provides no reasonable estimates for this data.

In the constant returns to scale model, the results are more favourable. All coefficients estimates are statistically significant but when four-period, two- and four-period, three- and four-period, and two-, three- and four-period lagged levels are used as instruments the coefficient estimates are greater than or close to one. With all other instrument sets we get reasonable estimates for the capital coefficient, and in the case of two-period and two- and three-period lagged levels the constant estimate is also statistically significant. As noted above, we wish to use the instrument set which satisfies the Sargan Test and which has lagged values a number of periods back. We find, in this case, that from amongst the instrument sets which yield reasonable coefficient estimates, the instrument set comprising two- and three-period lagged levels has a Sargan statistic with a reasonably high p-value. This instrument set is therefore considered appropriate. We also note that, for this instrument set, the χ^2 statistic is significantly higher in the constant returns to scale model than in the standard model. This suggests that the restriction to the model is appropriate.

Table D.12: Anderson-Hsaio Estimates for the water sector

Standard model $Z = (k, l)$	Lag2	Lag3	Lag4	Lag2+3	Lag2+4	Lag3+4	Lag 2+3+4
α	-..2799	-.2657	-.4835	-.2961	-.2387	-.2184	-.1758
	(1.385)	(.6539)	(.6279)	(.3599)	(.3328)	(.3342)	(.2709)
β	-.1422	-.0404	.0325	-.0646	-.0151	-.0189	-.0212
	(.3694)	(.1617)	(.1043)	(.1079)	(.0672)	(.0462)	(.0421)
a	.0070	.0088*	.0130*	.0087*	.0117*	.0116*	.0113*
	(.0072)	(.0025)	(.0048)	(.0022)	(.0031)	(.0032)	(.0029)
R^2(overall)	0.88	0.83	0.77	0.84	0.81	0.82	0.83
χ^2	8.40	14.51	12.23	12.5	16.17	16.19	16.23
S	n.a.	n.a.	n.a.	0.64	1.05	0.98	0.82
S p-value	n.a.	n.a.	n.a.	0.73	0.59	0.62	0.94
Observations	70	60	50	60	50	50	50
CRS model $Z = (k - l)$	**Lag2**	**Lag3**	**Lag4**	**Lag2+3**	**Lag2+4**	**Lag3+4**	**Lag 2+3+4**
α	0.8288*	.8772*	1.118*	.8642*	1.058*	1.010*	.9821*
	(.2712)	(.1605)	(.1633)	(.1352)	(.1338)	(.1072)	(.0999)
a	.0125*	.0120	.0019	.0119*	.0015	.0011	.0009
	(.0038)	(.0038)	(.0044)	(.0039)	(.0039)	(.0037)	(.0037)
R^2(overall)	0.41	0.42	0.44	0.42	0.44	0.44	0.44
χ^2	134.1	171.65	100.81	176.76	128.78	160.75	168.59
S	n.a.	n.a.	n.a.	0.74	0.91	0.93	0.94
S p-value	n.a.	n.a.	n.a.	0.39	0.35	0..34	0.63
Observations	70	60	50	60	50	50	50

We conclude that, given our dataset, the preferred model for the water sector is the constant returns to scale model. The coefficient on capital is estimated using the two- and three–period lagged values of $(k - l)$ *as instruments.* We explore the properties of the residual in this model to test for autocorrelation and heteroscedasticity.

Figure D.27 plots the squared Anderson Hsiao residuals (X-axis) against the fitted value of output (Y-axis). There seems to be a relationship between the squared residual and the fitted value, suggesting that the variance of the residual is not constant. Heteroscedasticity may therefore be present in this model. This does not affect the consistency of the estimates but the Sargan Test results may not be valid.

Figure D.28 plots the residuals (Y-axis) against their lagged values (X-axis). There is no distinct relationship between these variables. We also run the following regression:

$$\Delta\varepsilon_{it} = \rho\Delta\varepsilon_{it-1} + v_{it}$$

We find that there is only a small correlation between the error term and its lagged value in this model ($\widehat{\rho} = -0.13$). Autocorrelation is therefore unlikely to be a significant problem.

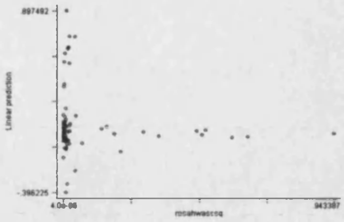

Figure D.27: Water and Anderson-Hsiao estimation: relationship between squared residuals and fitted values in the CRS model

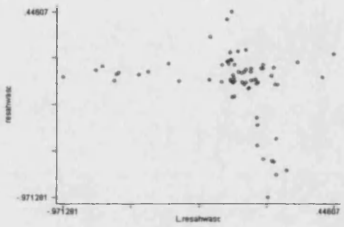

Figure D.28: Water and Anderson-Hsiao estimation: relationship between residuals and lagged residuals in the CRS model

D.5.3 Analysis

Our main conclusions on the Anderson-Hsiao estimates of our production functions are as follows.

- For both the electricity distribution and water sectors our preferred estimator arises when we focus on the constant returns to scale model with time dummies. In both sectors we use the instrument set comprising two- and three-period lagged levels of (k-l).

- The Anderson-Hsiao estimators are consistent if these instruments are uncorrelated with the error term, and if they are correlated with the current first-differences of capital and labour. The estimates are, however, potentially biased and they are inefficient.

- The capital and labour variables are highly persistent. The lagged levels may therefore provide weak instruments for the first-differenced regressors. This can result in substantial bias in the estimates. We have attempted to reduce this bias by using lagged levels further back in time as instruments.

- For both sectors there is evidence of potential heteroscedasticity in the error term. There is some loss of efficiency when the residuals are not homoscedastic. In addition, the Sargan Test may no longer be valid.
- There is no evidence of autocorrelation in either model.

D.6 System estimator

Blundell and Bond (1998) suggest using a generalized method of moments *system* estimator when weak instruments is a problem with the standard Anderson-Hsiao technique. This estimator includes lagged levels as instruments for the equation in first-differences, and lagged first-differences as instruments for the equation in levels. They assume that the error term (including the fixed-effect) is independent of the regressors in the model. The authors also impose first order serial correlation in the error term.

This estimator reduces the biases associated with the standard GMM estimator. The properties of the estimator are highly dependent on the panel data having a large N and a fixed T. As this does not apply here we do not explore this estimation method further. It is unlikely to provide improved results relative to Anderson-Hsiao estimation with our small sample size.

Appendix E

Impact of regulation: explanatory variables

We estimate, in section 4.6, the impact of different features of the regulatory regime on productivity growth in the water and electricity sectors between 1991 and 2000. The explanatory variables used in that regression are defined here.

E.1 Expected regulatory lag

As discussed in detail in Chapter 2, the firm's decision to undertake cost-reducing effort is determined by the expected profit which can be earned. The profit level depends on the length of time between when the firm makes the cost saving and when that saving is shared with consumers via price reductions. The longer this length of time - termed the regulatory lag - the higher the level of profit which the firm will earn from each unit of effort and, hence, the higher the effort level undertaken. Indeed, as discussed in Chapter 3, technical efficiency is maximised if the firm is allowed to retain these profits forever.

The rule used to share cost savings made in period t is set at the end of period t. The firm's effort decisions are therefore based on its expectation of what the regulatory lag will be. We assume that the firm basis its expectation on the methodology which was used at the most recent periodic review, updated for any recent announcements by the regulator. Table E.1 shows the expected regulatory lag for each sector and gives an explanation as to why the lag has changed over time. Our analysis of the firm's expectation is based on a review of the different rules used to share operating cost savings with consumers over time. These rules are discussed in more detail in section 2.2.1 and Appendix A.

The differences across sectors reflects the decisions made by individual regulators on the appropriate length of the lag, and on the preferred way to share savings over time. A key factor affecting this difference is whether the regulator used a large price cut at the start of the period, or whether savings were shared over the period through a glidepath adjustment to the annual X-factor. At privatisation there was no clear methodology about how savings would be shared with consumers. In particular, the government did not indicate at the start of the first regulatory period the point in time at which savings would be shared. We therefore assume that in the early years after privatisation, up to the point where the regulator first makes a clear statement on the matter, that the firm expects to keep savings made to the end of the next regulatory period. That is, if savings are made in the first year of period t, the firm expects that they will be shared with consumers in the last year of period t+1.

Table E.1: Expected regulatory lag (no of years)

	Lag	Explanation
Distribution		
1991-92	8	Expect savings to be shared at end of next period: 1990-00
1992-93	7	
1993-94	6	
1994-95	6	Expect savings to be shared in inital price cut in 2000-01.
1995-96	5	
1996-97	4	
1997-98	3	
1998-99	7	Expect savings to be shared in initial price cut in 2005-06.
1999-00	6	
NGC		
1991-92	4	Expect savings to be shared at end of next period: 1995-96
1992-93	4	Expect savings to be shared at end of next period: 1996-97
1993-94	7	Expect savings to be shared at end of next period: 2000-01
1994-95	6	
1995-96	2	See P_0 cut for RECs. Expect similar treatment.
		Expect savings to be shared in inital price cut in 1997/98
1996-97	5	Expect savings to be shared in inital price cut in 2001/02
1997-98	4	
1998-99	3	
1999-00	2	
Water		
1991-92	13	Expect savings to shared by end of next period: 2004/05
1992-93	12	
1993-94	11	
1994-95	10	
1995-96[1]	9	
1996-97[1]	8	
1997-98	3	Expect savings to be shared in initial price cut in 2000/01
1998-99	2	
1999-00	5	Expect savings to shared via the five-year rolling adjustment

Note: [1]In July 1995 the MMC set the price cap for South West Water Services. A five-year glidepath was used for sharing savings with customers, in contrast to the ten-year glidepath used by Ofwat for all other companies. South West Water's expected regulatory lag in 1995/96 is therefore 4 years and in 1996/97 it is 3 years. In all other years, South West Water has the same expected regulatory lag as the other WASCs.

E.2 Commitment

In theory, the contract agreed at the most recent periodic review is fixed for the duration of the regulatory period. The regulator is assumed to be committed to the contract and to the price-cap setting methodology used at the most recent periodic review. In practice, however, the regulator has the discretion to propose changes to the contract mid-period and to change the methodology used to set the price cap, including the rule used to share savings. This reduces the expected share of savings. It is therefore the interaction of this variable with the expected regulatory lag which is expected to affect productivity. The less committed the regulator is expected to be, the lower the effective regulatory lag (expected lag $\times$ expected commitment) and, hence, the lower the incentive to make cost savings.

The firm takes account of the risk that the regulator will change the detail of the contract when forming its expectation of the profit which it will earn from each unit of cost-reducing effort. As discussed in section 3.1.2, the firm will reduce its effort level if it believes that the regulator is less than 100% committed to the regulatory contract. This provides the firm with some insurance against the possibility that the actual profit earned will be less than expected.

We assume that the firm's belief is determined by the extent to which the regulator demonstrated that he was committed to recent contracts. If the regulator did not change the contract, or the 'rules of the game', significantly in the past, the firm will assume that the regulator remains committed to the contract. If, however, the regulator did make significant changes, the firm will assume that future changes are more likely and, hence, will expect that the regulator is not fully committed to the contract. In this way, the regulator's prior actions directly impact on the firm's belief about his commitment level and, hence, on the firm's effort choice. Similarly, if the government indicates that it is to reduce the profits earned from each unit of effort, through a 'windfall tax', the firm will assume that the contract is not fixed.

To capture this idea we construct a commitment index for each sector. This index reflects the firm's belief about how committed the regulator is to the fixed regulatory contract. We assume that the firm expects the regulator to be 100% committed to the contract at privatisation - indicated by an index of 1. Any announcements or decisions which indicate that the regulator, or the government, will share a higher proportion of savings with consumers, or will share them earlier, is taken as a signal that the regulator is not committed to the fixed-price contract[1]. A change which has a major impact on share prices is considered a major shock, and is reflected as a 0.25 reduction

[1]We assume that the firm's belief about the regulator's commitment is only affected by decisions made by its own regulator. In practice, the firm's expectation may also be affected by decisions made by the Competition Commission and other regulators.

in the index. A change which has less of an impact on share prices is considered a moderate shock, and reduces the index by 0.1. Other minor shocks, which have little, if any, impact on share prices, reduce the index by 0.05. Figures E.1 to E.3 show share price movements since privatisation. These are used to determine which shocks were moderate and which were extreme. We assume that there is a two-year lag between when an announcement is made and when the firm is able to adjust its cost decisions to reflect the new methodology. The commitment index in year t therefore reflects the firm's reaction to events which occurred in years t-2 and earlier.

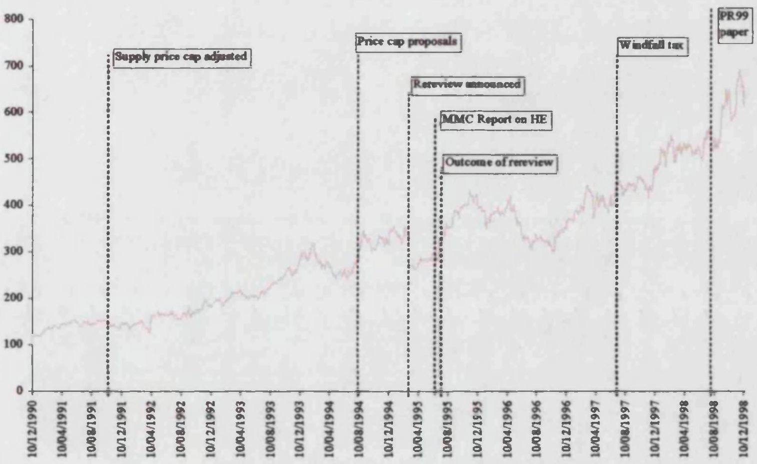

Figure E.1: The RECs share price index

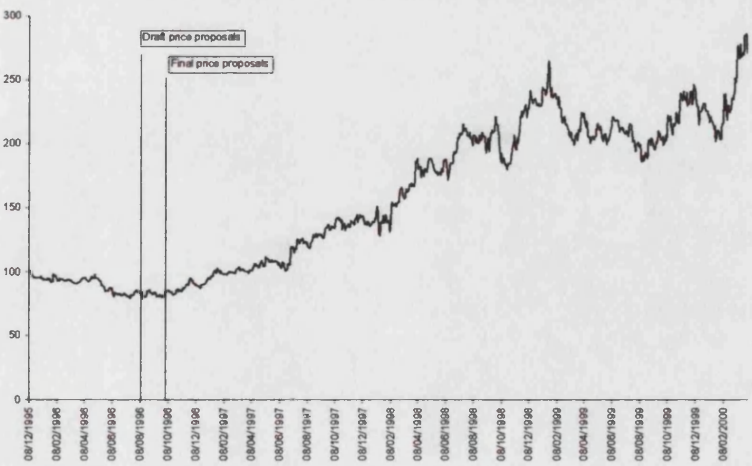

Figure E.2: NGC's share price index

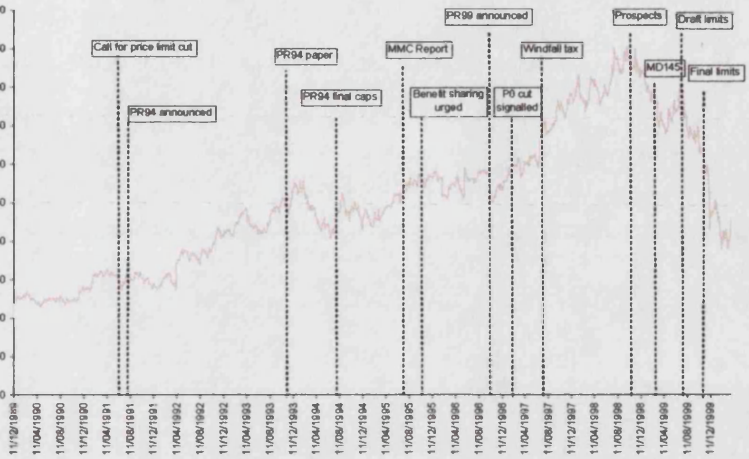

Figure E.3: WASCs share price index

Table E.2 presents the commitment index for the electricity distribution, electricity transmission and water sectors. The main contract changes which have been made by the regulators since privatisation are discussed in Appendix A. We show the value of the index in each year and provide a brief explanation of why its value has changed in particular years. We note that the regulator appears to have remained committed to NGC's regulatory contract in all years since privatisation. In addition, the transmission company did not have to pay the Government's windfall tax in 1997.

Table E.2: Commitment Index (1=full commitment)

Year	Index	Explanation
Distribution		
1991-92 to 1996/97	1	Full commitment
1997-98	0.75	24/03/95: rereview of price caps announced Regulator signals disregard for fixed contract
1998-99 & 1999-00	0.65	02/07/97: windfall tax announced Profits arising from efficiency savings removed
NGC		
1991-92 to 1999-00	1	Full commitment
Water		
1991-92 to 1994-95	1	Full commitment
1995-96 to 1997-98	0.95	02/10/92: Price caps adjusted for difference between actual and expected COPI
1998-99	0.85	02/07/97: windfall tax announced Profits arising from efficiency savings removed
1999-00	0.60	26/06/97: Initial PR99 proposals stress use of P_0 cuts. Change in methodology on sharing of efficiency savings Move from 10-year to 5-year glidepath

E.3 Years since privatisation

It is assumed that at privatisation there was a step-change in productivity, reflecting the impact of the change in ownership. A key issue with any empirical investigation in the utility sectors is that regulation and privatisation were introduced at the same time. It is therefore difficult to separate the 'privatisation effect' from the 'regulation effect'. This variable - the number of years since privatisation - is intended to control for the 'privatisation effect' to some extent.

If privatisation is the main factor affecting cost savings and productivity, we expect the rate of improvement to decline the longer the time period since privatisation. In contrast, if regulation has an ongoing and significant impact, cost savings and productivity improvement would continue to be strong as the private firm matures. The presumption is that productivity growth later in the period would be attributed to the impact of regulation rather than privatisation.

The relevant privatisation dates for each of the sectors are as follows:

- Electricity distribution and NGC: 31/03/90;
- Water: 31/12/89.

As always, the year of interest is the financial year running from April 1st to March

31st. For the electricity companies, 1991/92 is considered to be 1 year after privatisation. For the water companies, it is considered to be 1.33 years after privatisation.

E.4 Corporate activity

A firm's cost saving potential, and productivity growth rate, are expected to improve if the firm is taken-over and/or if it merges with another company. Indeed this is a common rational for such corporate activity. In addition, when regulatory approval is given for a takeover or merger, the regulator generally requires the firm to deliver higher than previously assumed cost savings. We therefore expect that a takeover or merger in the previous year will lead to cost reductions in the next year. This is not guaranteed, however. Vickers and Yarrow (1988a) analyse the impact of a takeover, or a takeover threat, on the firm's performance. They conclude that the impact is not always positive and that it 'depends heavily upon the precise characteristics of the relevant capital markets'. There is therefore a possibility that corporate activity will have little impact on productivity.

We examine this issue by constructing a dummy variable which is 0 in year t if there was no corporate activity in year t-1, and 1 if there was a merger or takeover in year t-1. In the case of NGC we also consider the flotation of the company on the London Stock Exchange in 1995 as a significant change in corporate structure which is likely to affect productivity. We note that the corporate activity relates to the Group company, and the impact on the regulated business may not be significant. In addition, the lag between the corporate activity taking place and productivity improvements being delivered may be longer than one year.

E.5 Output and quality of service

Finally, we introduce a scale variable to reflect the fact that the level of output, and the quality of supply delivered, will have an impact on the change in productivity. The variable used is the quality-adjusted output index, described in Appendix C. We expect that, for a given set of inputs, the higher the level of quality-adjusted output delivered the higher the rate of change in productivity. Where the outputs and inputs both change the impact on TFP growth is not immediately evident as it depends on the relative rate of change in these variables.

Appendix F

Comparing delegation contracts

F.1 Properties of the welfare function

The level of welfare, incorporating the transfer restriction, is:

$$W(q,\theta) = \frac{(1-2\lambda)\,q^2}{2b} + \frac{\lambda\,(a-b\theta)\,q}{b}$$

Lemma F.1 *The level of welfare is increasing in output for a firm of type θ.*

Proof. The impact of a change in output on welfare is measured as:

$$\frac{\partial W(q,\theta)}{\partial q} = \frac{(1-2\lambda)q + \lambda\,(a-b\theta)}{b} \tag{F.1}$$

An extremum of the welfare function exists at the point where this derivative is zero:

$$\begin{aligned}
\frac{(1-2\lambda)q + \lambda\,(a-b\theta)}{b} &= 0 \\
(2\lambda-1)q &= \lambda\,(a-b\theta) \\
q(\theta) &= \left(\frac{\lambda}{2\lambda-1}\right)(a-b\theta)
\end{aligned}$$

To determine whether welfare is maximised or minimised at this level of output we must check the second-order conditions. We have:

$$\frac{\partial^2 W(q,\theta)}{\partial q^2} = \frac{1-2\lambda}{b} \tag{F.2}$$

For a convex welfare function ($\lambda < \frac{1}{2}$) we have:

$$\begin{aligned}
\frac{\partial^2 W(\theta)}{\partial q^2} &> 0 \\
&\Rightarrow\ q(\theta) = \left(\frac{\lambda}{2\lambda-1}\right)(a-b\theta) \text{ is a minimum}
\end{aligned}$$

For a concave welfare function ($\lambda > \frac{1}{2}$) we have:

$$\begin{aligned} \frac{\partial^2 W(\theta)}{\partial q} &< 0 \\ &\Rightarrow q(\theta) = \left(\frac{\lambda}{2\lambda - 1}\right)(a - b\theta) \text{ is a maximum} \end{aligned}$$

Consider first the case of a convex welfare function
The welfare function is minimised at $q(\theta) = \left(\frac{\lambda}{2\lambda-1}\right)(a - b\theta)$.
This implies that:

$$\begin{aligned} \frac{\partial W(\theta)}{\partial q} &< 0 \text{ if } q(\theta) < \left(\frac{\lambda}{2\lambda - 1}\right)(a - b\theta) \\ \frac{\partial W(\theta)}{\partial q} &> 0 \text{ if } q(\theta) > \left(\frac{\lambda}{2\lambda - 1}\right)(a - b\theta) \end{aligned}$$

$q(\theta) < \left(\frac{\lambda}{2\lambda-1}\right)(a - b\theta)$ is not feasible here as $\left(\frac{\lambda}{2\lambda-1}\right) < 0$ and we require $q(\theta) > 0$
.

The only feasible values of $q(\theta)$ therefore result in:

$$\frac{\partial W(\theta)}{\partial q} > 0$$

Next consider the case of a concave welfare function
The welfare function is maximised at $q(\theta) = \left(\frac{\lambda}{2\lambda-1}\right)(a - b\theta)$.
This implies that:

$$\begin{aligned} \frac{\partial W(\theta)}{\partial q} &> 0 \text{ if } q(\theta) < \left(\frac{\lambda}{2\lambda - 1}\right)(a - b\theta) \\ \frac{\partial W(\theta)}{\partial q} &< 0 \text{ if } q(\theta) > \left(\frac{\lambda}{2\lambda - 1}\right)(a - b\theta) \end{aligned}$$

$q(\theta) > \left(\frac{\lambda}{2\lambda-1}\right)(a - b\theta)$ is not feasible here as $\left(\frac{\lambda}{2\lambda-1}\right) > 1$ and we require $q(\theta) < a - b\theta$.

The only feasible values of $q(\theta)$ therefore result in:

$$\frac{\partial W(\theta)}{\partial q} > 0$$

We have thus shown that, for all values of λ and for all feasible values of output, welfare is increasing with the level of output.

■

Lemma F.2 *The rate of change in welfare, arising from a change in output, is increasing in the level of output if the welfare function is convex.*

Proof. The impact of the level of output on the rate of change in welfare is calculated as:

$$\frac{\partial^2 W(q,\theta)}{\partial q^2} = \frac{1-2\lambda}{b}$$

If the welfare function is convex ($\lambda \in [0, \frac{1}{2})$) the function $\frac{1-2\lambda}{b}$ is non-negative. The rate of change in welfare is therefore increasing with the level of output. ■

Lemma F.3 *The rate of change in welfare, arising from a change in output, is decreasing in the level of output if the welfare function is concave.*

Proof. The impact of the level of output on the rate of change in welfare is calculated as:

$$\frac{\partial^2 W(q,\theta)}{\partial q^2} = \frac{1-2\lambda}{b}$$

If the welfare function is concave ($\lambda \in (\frac{1}{2}, 1]$) the function $\frac{1-2\lambda}{b}$ is non-positive. The rate of change in welfare is therefore decreasing with the level of output. ■

Lemma F.4 *Welfare is decreasing in type for a given level of output. The rate of decrease is constant for all types.*

Proof. The impact of a change in type on the level of welfare is:

$$\frac{\partial W(q,\theta)}{\partial \theta} = \left(\frac{1-2\lambda}{b}\right) q(\theta) \left(\frac{\partial q(\theta)}{\partial \theta}\right) + \frac{\lambda\,(a-b\theta)\left(\frac{\partial q(\theta)}{\partial \theta}\right)}{b} - \lambda q(\theta) \qquad \text{(F.3)}$$

Let the output level be constant for all types. We therefore have $\left(\frac{\partial q(\theta)}{\partial \theta}\right) = 0$ and the derivative becomes:

$$\frac{\partial W(q,\theta)}{\partial \theta} \mid_{q(\theta)=q} = -\lambda q$$

We therefore see that for a given output level an increase in type reduces welfare and the rate of decrease is the same for all types. ■

Lemma F.5 *The rate of change in welfare is decreasing in type for a given level of output. The rate of decrease is constant for all types.*

Proof. The impact of a change in the firm's type on the rate of change in welfare is:

$$\frac{\partial W(q,\theta)}{\partial q \partial \theta} = \frac{(1-2\lambda)\left(\frac{\partial q(\theta)}{\partial \theta}\right) - b\lambda}{b} \tag{F.4}$$

Let the output level be constant for all types. We therefore have $\left(\frac{\partial q(\theta)}{\partial \theta}\right) = 0$ and the derivative becomes:

$$\frac{\partial W(q,\theta)}{\partial q \partial \theta} \Big|_{q(\theta)=q} = -\lambda$$

The rate of change in welfare is therefore decreasing in type for a given output level. The rate of decrease equal to the weight placed on profits in the welfare function. ∎

F.2 Output choices under Contract R^2

Under a delegation contract the firm chooses its monopoly level of output, $q^M(\theta)$, if it lies within the delegated choice set. If not, the firm chooses the nearest level of output as the profit function is concave and symmetric in output.

Say the regulator allows the firm to choose its output level from the restricted two interval choice set $[q_{\min}, X+k] \cup [Y-k, \infty]$. The firm's optimal output choices are derived below. The results are summarised in Table F.1.

- $q^M(\theta) \leq q_{\min}$: As the monopoly level of output is not available the firm chooses the nearest available level of output which is $q_{\min}$.
- $q_{\min} < q^M(\theta) \leq X+k$: The monopoly level of output is available and hence it is chosen by the firm.
- $X+k < q^M(\theta) \leq \frac{X+Y}{2}$: As the monopoly level of output is not available the firm chooses the nearest available level of output which is $X+k$.
- $\frac{X+Y}{2} < q^M(\theta) \leq Y-k$: As the monopoly level of output is not available the firm chooses the nearest available level of output which is $Y-k$.
- $Y-k < q^M(\theta)$: The monopoly level of output is available and hence it is chosen by the firm.

Table F.2 provides the optimal output choices as a function of the firm's type. The regions are calculated using the fact that $q^M(\theta) = \frac{a-b\theta}{2}$. The calculations are as follows:

- $q^M(\theta) \leq q_{\min} \Rightarrow \frac{a-b\theta}{2} \leq a - b\theta_H \Rightarrow \frac{a-2(a-b\theta_H)}{b} \leq \theta$
- $q_{\min} < q^M(\theta) \leq X+k \Rightarrow a-b\theta_H < \frac{a-b\theta}{2} \leq X+k \Rightarrow \frac{a-2(X+k)}{b} \leq \theta < \frac{a-2(a-b\theta_H)}{b}$

- $X+k < q^M(\theta) \le \frac{X+Y}{2} \Rightarrow X+k < \frac{a-b\theta}{2} \le \frac{X+Y}{2} \Rightarrow \frac{a-(X+Y)}{b} \le \theta < \frac{a-2(X+k)}{b}$
- $\frac{X+Y}{2} < q^M(\theta) \le Y-k \Rightarrow \frac{X+Y}{2} < \frac{a-b\theta}{2} \le Y-k \Rightarrow \frac{a-2(Y-k)}{b} \le \theta < \frac{a-(X+Y)}{b}$
- $Y-k < q^M(\theta) \Rightarrow Y-k < \frac{a-b\theta}{2} \Rightarrow \theta < \frac{a-2(Y-k)}{b}$

Table F.2: Output Choices by Type

Type region	Choice
$\theta_L \le \theta < \frac{a-2(Y-k)}{b}$	$q^M(\theta)$
$\frac{a-2(Y-k)}{b} \le \theta < \frac{a-(X+Y)}{b}$	$Y-k$
$\frac{a-(X+Y)}{b} \le \theta < \frac{a-2(X+k)}{b}$	$X+k$
$\frac{a-2(X+k)}{b} \le \theta < \frac{a-2q_{\min}}{b}$	$q^M(\theta)$
$\frac{a-2q_{\min}}{b} \le \theta < \theta_H$	$q_{\min}$

F.3 Restricted delegation with a uniform distribution

We consider here the regulator's preference between the output floor contract and the restricted two-interval contract when the firm's type is assumed to be distributed uniformly on the interval $[\theta_L, \theta_H]$.

Proposition F.1 *If the firm's type is distributed uniformly on the interval $[\theta_L, \theta_H]$ the output floor contract is not best.*

Proof. Consider a restricted two-interval contract which allows the firm to choose its level of output from the set:

$$[q_{\min}, X+k] \cup [Y-k, \infty]$$

When the firm's type is distributed uniformly on the interval $[\theta_L, \theta_H]$ the change in expected welfare, arising from a change in k , is:

$$\Delta = \frac{1}{\theta_H - \theta_L}\left\{\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} d\theta - \int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} d\theta\right\}$$

Table F.1: Output Choices

Monopoly output level	Choice with choice set Q_i
$q^M(\theta) \le q_{\min}$	$q_{\min}$
$q_{\min} < q^M(\theta) \le X+k$	$q^M(\theta)$
$X+k < q^M(\theta) \le \frac{X+Y}{2}$	$X+k$
$\frac{X+Y}{2} < q^M(\theta) \le Y-k$	$Y-k$
$Y-k < q^M(\theta)$	$q^M(\theta)$

$$= \frac{1}{\theta_H - \theta_L} \left\{ \begin{array}{c} \int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \left[\frac{(1-2\lambda)(X+k)+\lambda(a-b\theta)}{b}\right] d\theta \\ -\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \left[\frac{(1-2\lambda)(Y-k)+\lambda(a-b\theta)}{b}\right] d\theta \end{array} \right\}$$

The first integral inside the brackets is calculated as:

$$\begin{aligned}
&\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \left[\frac{(1-2\lambda)(X+k)+\lambda\,(a-b\theta)}{b}\right] d\theta \\
&= \left\{ \begin{array}{c} \left[\frac{(1-2\lambda)(X+k)+a\lambda}{b}\right] \left[\left(\frac{a-2(X+k)}{b}\right) - \left(\frac{a-(X+Y)}{b}\right)\right] \\ -\left[\frac{\lambda}{2}\right] \left[\left(\frac{a-2(X+k)}{b}\right)^2 - \left(\frac{a-(X+Y)}{b}\right)^2\right] \end{array} \right\} \\
&= \left\{ \begin{array}{c} \left[\frac{(1-2\lambda)(X+k)+a\lambda}{b}\right] \left(\frac{Y-X-2k}{b}\right) \\ -\left[\frac{\lambda}{2}\right] \left(\frac{Y-X-2k}{b}\right) \left(\frac{2a-3X-Y+2k}{b}\right) \end{array} \right\} \\
&= \left(\frac{Y-X-2k}{b}\right)\left(\frac{X+k}{b}\right) + \left(\frac{\lambda}{2}\right)\left(\frac{Y-X-2k}{b}\right)^2
\end{aligned}$$

Similarly, the second integral is calculated as:

$$\begin{aligned}
&\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \left[\frac{(1-2\lambda)\,(Y-k)+\lambda\,(a-b\theta)}{b}\right] d\theta \\
&= \left(\frac{Y-X-2k}{b}\right)\left(\frac{Y-k}{b}\right) - \left(\frac{\lambda}{2}\right)\left(\frac{Y-X-2k}{b}\right)^2
\end{aligned}$$

The change in constrained expected welfare is therefore equal to:

$$\begin{aligned}
\Delta &= \frac{1}{\theta_H - \theta_L} \left\{ \begin{array}{c} \left(\frac{Y-X-2k}{b}\right)\left(\frac{X+k}{b}\right) + \left(\frac{\lambda}{2}\right)\left(\frac{Y-X-2k}{b}\right)^2 \\ -\left(\frac{Y-X-2k}{b}\right)\left(\frac{Y-k}{b}\right) + \left(\frac{\lambda}{2}\right)\left(\frac{Y-X-2k}{b}\right)^2 \end{array} \right\} \\
&= \frac{1}{\theta_H - \theta_L} \left\{\lambda\left(\frac{Y-X-2k}{b}\right)^2 - \left(\frac{Y-X-2k}{b}\right)^2\right\}
\end{aligned}$$

Given $0 < \lambda < 1$ this function is negative **for any** X, Y, k. We therefore conclude that as we increase the value of k (reduce the size of the interval of disallowed outputs) welfare decreases. The regulator is therefore better off with a more restricted contract (smaller k).

The output floor contract has the maximum k and hence we see that it will always yield a lower level of welfare than a restricted two-interval contract. That is, the output floor contract is not optimal when the regulator believes that the firm's type is uniformly distributed on the interval $[\theta_L, \theta_H]$. ■

F.4 Optimality of the restricted two-interval contract

Notation

$\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta$ is Integral X

$\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta$ is Integral Y

$\frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)$ is Integrand X

$\frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)$ is Integrand Y

I_1 is the region of Types $\left[\frac{a-2(Y-k)}{b}, \frac{a-(X+Y)}{b}\right]$

I_2 is the region of Types $\left[\frac{a-(X+Y)}{b}, \frac{a-2(X+k)}{b}\right]$

Lemma F.6 *(Lemma 5.1) Suppose locally the regulator's belief function, $f(\theta)$, is non-increasing and concave and the welfare function is concave ($\lambda \epsilon \left(\frac{1}{2}, 1\right)$). Then welfare from the restricted two-interval choice set increases with the size of the interval of disallowed output levels. That is, the more restriction there is in the two-interval choice set (ie, the lower is k) the higher is the level of welfare.*

Proof. Step 1: General conditions for welfare to decrease as k increases

Consider the restricted two-interval choice set which allows the firm to choose its level of output from the set:

$$[q_{\min}, X+k] \cup [Y-k, \infty]$$

The smaller the value of k, the wider is the set of disallowed choices and, hence, the more restriction is placed on the firm's decision.

The rate of change in expected welfare from an increase in k is:

$$\Delta = \int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta - \int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta$$

If an increase in k leads to a decrease in welfare **for all feasible** k ($\Delta < 0$), then a two-interval choice set with a large interval of disallowed output levels (ie, small k) will yield a higher level of welfare than a two-interval choice set with a smaller interval of disallowed output levels (ie, large k).

With a concave welfare function ($\lambda > \frac{1}{2}$) the rate of change in welfare from a change in output is decreasing in the level of output for any given type (Lemma F.3). We therefore have:

$$\frac{\partial W(Y-k,\theta)}{\partial q(\theta)} < \frac{\partial W(X+k,\theta)}{\partial q(\theta)} \textbf{ for any given } \boldsymbol{\theta}$$

$$\Rightarrow \text{Integrand Y} < \text{Integrand X } \textbf{for any given } \boldsymbol{\theta}$$

Given this, the sufficient conditions for $\Delta < 0$ (Integral Y>Integral X) are as follows.

1. Integral Y starts at a higher value than Integral X.

2. Both Integrands are decreasing in type.

3. Because Integrand X lies above Integrand Y, we require the slope of Integrand X in the region I_2 to be steeper (ie, more negative) than the slope of Integrand Y in the region I_1. This will occur if both functions are concave in type.

We show that the conditions outlined in this lemma are **sufficient** to ensure that these conditions are met. Figure F.1 illustrates functions which satisfy these conditions.

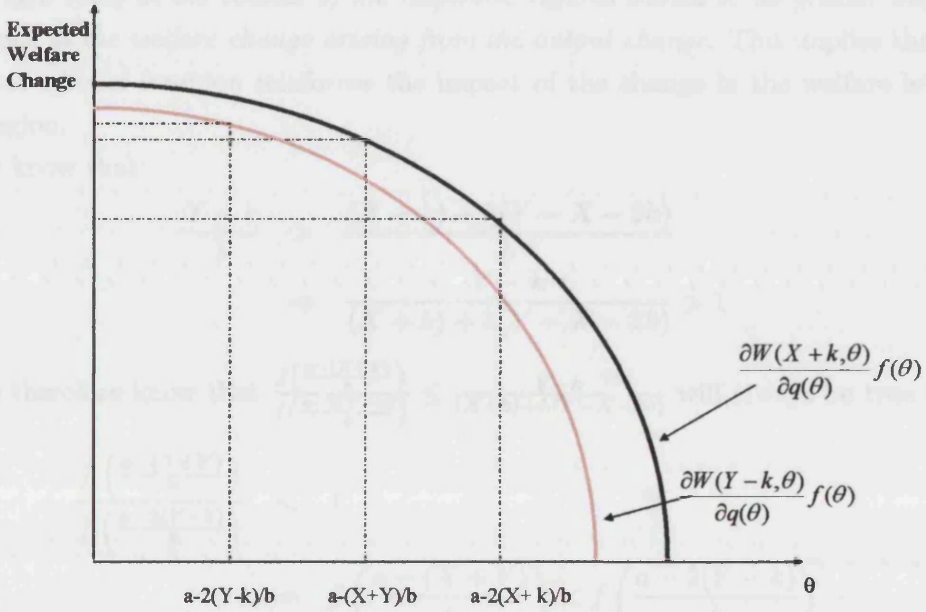

Figure F.1: Welfare change with a decreasing and concave density function

Step 2: Calculating the starting point of the integrals

The value of Integrand Y at the left-hand boundary of its interval of integration is:

$$\frac{\partial W(Y-k, \frac{a-2(Y-k)}{b})}{\partial q(\theta)} = \left(\frac{Y-k}{b}\right) f\left(\frac{a-2(Y-k)}{b}\right)$$

The value of Integrand X at the left-hand boundary of its interval of integration is:

$$\frac{\partial W(X+k, \frac{a-(X+Y)}{b})}{\partial q(\theta)} = \left(\frac{(X+k)+\lambda(Y-X-2k)}{b}\right) f\left(\frac{a-(X+Y)}{b}\right)$$

Integral Y therefore starts at a higher value than Integral X if:

$$\left(\frac{Y-k}{b}\right) f\left(\frac{a-2(Y-k)}{b}\right) \geq \left(\frac{(X+k)+\lambda(Y-X-2k)}{b}\right) f\left(\frac{a-(X+Y)}{b}\right)$$

$$\frac{f\left(\frac{a-(X+Y)}{b}\right)}{f\left(\frac{a-2(Y-k)}{b}\right)} \leq \frac{Y-k}{(X+k)+\lambda(Y-X-2k)}$$

That is, the ratio of the proportion of types at the extremes of the region I_1 should be no greater than the ratio of the rate of change in welfare from the change in output at these extremes. Alternatively, we could say that *the change in the probability of the firm's type being at the bounds of the respective regions should be no greater than the difference in the welfare change arising from the output change.* This implies that the regulator's belief function reinforces the impact of the change in the welfare level in each region.

We know that:

$$\frac{Y-k}{b} > \frac{(X+k)+\lambda(Y-X-2k)}{b}$$
$$\Rightarrow \frac{Y-k}{(X+k)+\lambda(Y-X-2k)} > 1$$

We therefore know that $\frac{f\left(\frac{a-(X+Y)}{b}\right)}{f\left(\frac{a-2(Y-k)}{b}\right)} \leq \frac{Y-k}{(X+k)+\lambda(Y-X-2k)}$ will always be true if :

$$\frac{f\left(\frac{a-(X+Y)}{b}\right)}{f\left(\frac{a-2(Y-k)}{b}\right)} < 1$$
$$\Rightarrow f\left(\frac{a-(X+Y)}{b}\right) < f\left(\frac{a-2(Y-k)}{b}\right)$$

A sufficient condition for Integral Y to start at a higher point than Integral X is therefore:

$$\frac{\partial f(\theta)}{\partial \theta} < 1 \; \forall \; \theta$$

As an aside we note that this condition is always met with a uniform distribution since $\frac{\partial f(\theta)}{\partial \theta} = 0$.

Step 3: Calculating the slopes

The impact of a change in type on each of the integrands is calculated as:

$$\frac{\partial}{\partial \theta}\,[\text{Integrand}] = \frac{\partial W(s,\theta)}{\partial q(\theta)} \frac{\partial f(\theta)}{\partial \theta} - \lambda f(\theta)$$

for $s = Y-k\,, X+k$.

With $\frac{\partial W(s,\theta)}{\partial q(\theta)} > 0$, $\lambda \geq 0$ and $f(\theta) \geq 0$ a sufficient condition for the integrands to be decreasing is:

$$\frac{\partial f(\theta)}{\partial \theta} \leq 0$$

That is, *the density function should be non-increasing with type.*

This condition also ensures that $\frac{\partial f(\theta)}{\partial \theta} < 1$ and, hence, that Integral Y starts at a higher point than Integral X.

Step 4: Comparing the slopes

With the concave welfare function we have

$$\text{Integrand Y} < \text{Integrand X } \textbf{for any given } \boldsymbol{\theta}$$

As Integrand Y lies below Integrand X, for any given type, we need to compare the slopes of these functions to ensure that Integral Y is greater than Integral X.

We know that if we have two decreasing functions, A and B, where the slope of A is flatter than the slope of B, the Integral of A starts a higher point than the Integral of B, and the interval of integration is the same length for A and B, then the Integral of A will be greater than the Integral of B.

When the belief function is non-increasing in type Integral Y starts at a higher value than X, and the rate of change of the integrand is decreasing in type for $s = Y - k$ and $X + k$. Integral Y is therefore larger than Integral X if the slope of Integrand Y over region I_1 is less negative than the slope of Integrand X over I_2. This is true if the functions are concave.

The rate of change in the slope, for $s = Y - k$, $X + k$, is:

$$\frac{\partial^2}{\partial \theta^2}[\text{Integrand}] = \frac{\partial W(s,\theta)}{\partial q(\theta)} \frac{\partial^2 f(\theta)}{\partial \theta^2} - 2\lambda \frac{\partial f(\theta)}{\partial \theta}$$

We know that $\frac{\partial W(s,\theta)}{\partial \theta}$ is non-increasing (Lemma F.4). The derivative with respect to type must therefore be positive for the function to be concave (ie, with a concave non-increasing function the slope becomes 'more negative' as the type increases).

Given $\frac{\partial f(\theta)}{\partial \theta} \leq 0$, $\lambda \geq 0$ and $\frac{\partial W(s,\theta)}{\partial q(\theta)} \geq 0$ we have $\frac{\partial^2}{\partial \theta^2}\left[\frac{W(s,\theta)}{\partial q(\theta)} f(\theta)\right] \geq 0$ if:

$$\frac{\partial^2 f(\theta)}{\partial \theta^2} \geq 0$$

This occurs if the non-increasing belief function is concave.

Step 5: Comparing the integrals

If the welfare function is concave, and the regulator's belief function is non-increasing and concave, we have shown that:

1. Integral Y starts at a higher point than Integral X;

2. Both Integrands are decreasing in type; and

3. The slope of Integrand Y in I_1 is flatter than the slope of Integrand X in I_2 .

Combined these conditions allow us to conclude that the change in welfare from a change in k is non-positive. That is:

$$\left|\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right| \geq \left|\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right|$$
$$\Rightarrow \Delta \leq 0$$

Step 6: Conclusion

If the conditions of the Lemma hold an increase in k leads to a decrease in welfare for **any given** X, Y, k. The more restriction there is in the two-interval choice set (ie, the lower is k) the higher is the level of welfare. ∎

Lemma F.7 *(Lemma 5.1) Suppose locally the regulator's belief function, $f(\theta)$, is non-increasing and the welfare function is convex ($\lambda \epsilon \left(0, \frac{1}{2}\right)$). Then welfare from the restricted two-interval choice set increases with the size of the interval of disallowed output levels. That is, the more restriction there is in the two-interval choice set (ie, the lower is k) the higher is the level of welfare.*

Proof. This proof is exactly the same as the proof to Lemma F.6 except that Step 4 is no longer required.

With a convex welfare function ($\lambda < \frac{1}{2}$) the rate of change in welfare from a change in output is increasing in the level of output for any given type (Lemma F.2). We therefore have:

$$\frac{\partial W(Y-k,\theta)}{\partial q(\theta)} > \frac{\partial W(X+k,\theta)}{\partial q(\theta)}$$
$$\Rightarrow \text{Integrand Y} > \text{Integrand X } \textbf{for any given } \boldsymbol{\theta}$$

Given this, the sufficient conditions for $\Delta < 0$ (Integral Y>Integral X) are simply.

1. Integral Y starts at a higher value than Integral X.

2. Both Integrands are decreasing in type.

Steps 2, 3 and 5 in the proof to Lemma F.6 shows that if the regulator's belief function is non-increasing these sufficient conditions will hold. Figure F.2 illustrates functions which satisfy these conditions.

We therefore find that if the regulator's belief function is non-increasing, and the welfare function is convex:

$$\left|\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right| \geq \left|\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right|$$
$$\Rightarrow \Delta \leq 0$$

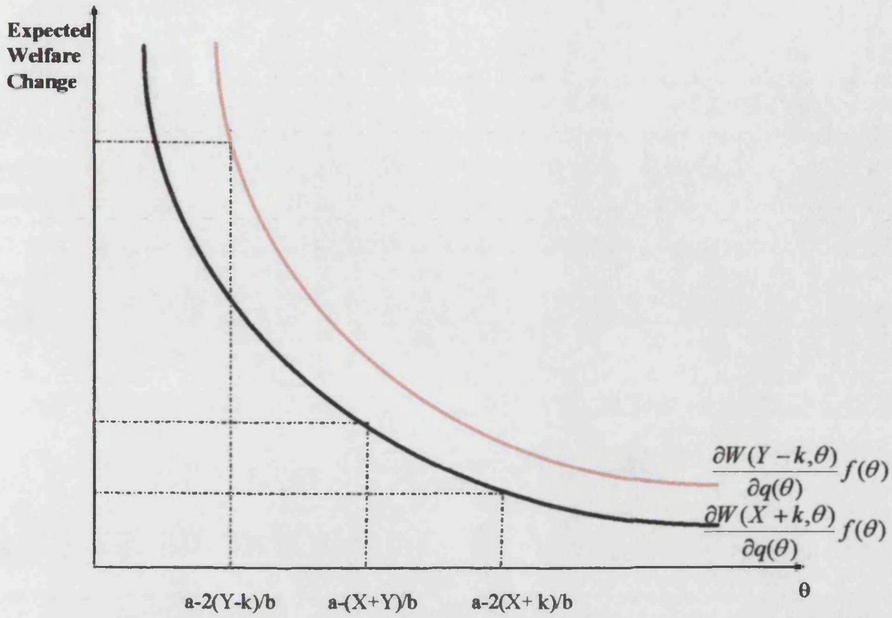

Figure F.2: Welfare change with decreasing density function

That is, an increase in k leads to a decrease in welfare for **any given** X, Y, k. The more restriction there is in the two-interval choice set (ie, the lower is k) the higher is the level of welfare. ■

F.5 Optimality of the output floor contract

Lemma F.8 *(Lemma 5.2) Suppose locally:*

(i) the regulator's belief function, $f(\theta)$, is increasing and convex;

(ii) the rate of increase is large enough to offset the decreasing rate of change in welfare:

$$\frac{\partial f(\theta)}{\partial \theta} \geq \frac{Y}{X} > 1;$$

(iii) the degree of curvature of the belief function is greater than a positive integer:

$$\left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) > \frac{b}{X};$$

(iv) the degree of convexity is sufficiently large:

$$\left(\frac{\partial^2 f(\theta)}{\partial \theta^2}\right)\left(\frac{1}{\frac{\partial f(\theta)}{\partial \theta}}\right) > \frac{2b\lambda}{X(1-2\lambda)} > 0; and$$

(v) the welfare function is convex ($\lambda \epsilon \left(0, \frac{1}{2}\right)$).

Then welfare from a restricted two-interval choice set decreases with the size of the interval of disallowed output levels. That is, the more restriction there is in the two-interval choice set (ie, the lower is k) the lower is the level of welfare.

Proof. Step 1: General conditions for welfare to increase as k increases

Consider the restricted two- interval choice set which allows the firm to choose its level of output from the set:

$$[q_{\min}, X + k] \cup [Y - k, \infty]$$

The smaller the value of k, the larger is the set of disallowed choices and, hence, the more restriction is placed on the firm's decision.

The rate of change in expected welfare from an increase in k is:

$$\Delta = \int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta - \int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta$$

If an increase in k leads to a increase in welfare **for all feasible** k ($\Delta > 0$), the two-interval choice set with a small interval of disallowed output levels (ie, large k) will yield a higher level of welfare than a two-interval choice set with a larger interval of disallowed output levels (ie, small k).

With a convex welfare function ($\lambda < \frac{1}{2}$) the rate of change in welfare from a change in output is increasing in the level of output for any given type (Lemma F.2). We therefore have:

$$\begin{aligned} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} &> \frac{\partial W(X+k,\theta)}{\partial q(\theta)} \textbf{ for any given } \boldsymbol{\theta} \\ &\Rightarrow \text{ Integrand Y} > \text{Integrand X } \textbf{for any given } \boldsymbol{\theta} \end{aligned}$$

Given this, the sufficient conditions for $\Delta > 0$ (Integral X>Integral Y) are as follows.

1. Integral X starts at a higher value than Integral Y.

2. Both Integrands are increasing in type.

3. Because Integrand X lies below Integrand Y, we require the slope of Integrand Y in the region I_1 to be flatter (ie, less positive) than the slope of Integrand X in the region I_2. This will occur if both functions are convex in type.

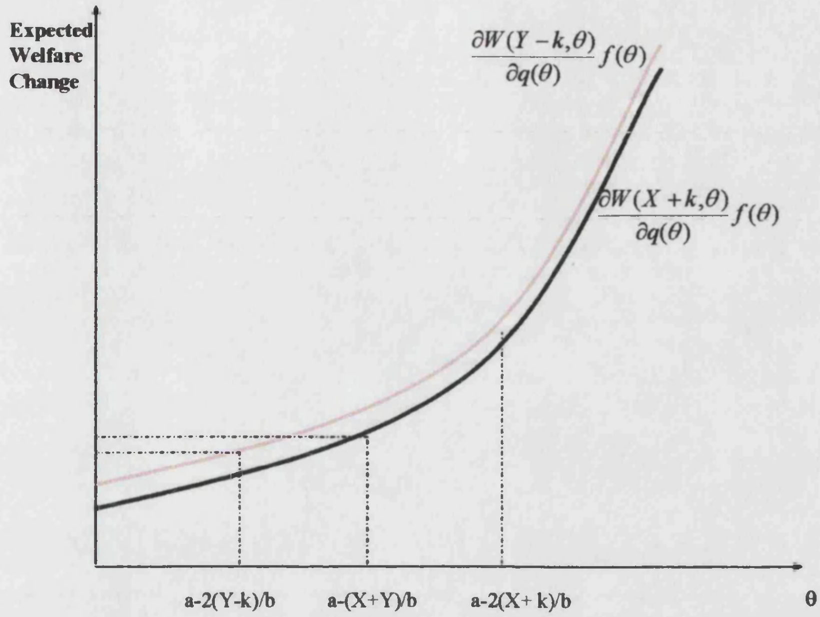

Figure F.3: Welfare change with an increasing and convex density function

We show that the conditions outlined in this lemma are **sufficient** to ensure that these conditions are met. Figure F.3 illustrates functions which satisfy these conditions.

Step 2: Calculating the starting point of the integrals

The value of Integrand Y at the left-hand boundary of its interval of integration is:

$$\frac{\partial W((Y-k,\frac{a-2(Y-k)}{b})}{\partial q(\theta)} = \left(\frac{Y-k}{b}\right) f\left(\frac{a-2(Y-k)}{b}\right)$$

The value of Integrand X at the left-hand boundary of its interval of integration is:

$$\frac{\partial W(X+k,\frac{a-(X+Y)}{b})}{\partial q(\theta)} = \left(\frac{(X+k)+\lambda(Y-X-2k)}{b}\right) f\left(\frac{a-(X+Y)}{b}\right)$$

Integral X therefore starts at a higher value than Integral Y if:

$$\left(\frac{(X+k)+\lambda(Y-X-2k)}{b}\right) f\left(\frac{a-(X+Y)}{b}\right) \geq \left(\frac{Y-k}{b}\right) f\left(\frac{a-2(Y-k)}{b}\right)$$

$$\frac{f\left(\frac{a-(X+Y)}{b}\right)}{f\left(\frac{a-2(Y-k)}{b}\right)} \geq \frac{Y-k}{(X+k)+\lambda(Y-X-2k)}$$

That is, the ratio of the proportion of types at the extremes of the region I_1 should be no less than the ratio of the rate of change in welfare from the change in output at these extremes. Alternatively, we say that *the change in the proportion of types over* *the bounds of the regions should be no less than the difference in the welfare change arising from the output change.*

A sufficient condition for $\frac{f\left(\frac{a-(X+Y)}{b}\right)}{f\left(\frac{a-2(Y-k)}{b}\right)} \geq \frac{Y-k}{(X+k)+\lambda(Y-X-2k)}$ is:

$$\frac{\partial f(\theta)}{\partial \theta} \geq \frac{Y-k}{(X+k)+\lambda(Y-X-2k)} \ \forall \theta$$

Taking account of all feasible values of $\lambda \epsilon \left(0, \frac{1}{2}\right)$ we know that:

$$\frac{Y-k}{(X+k)} > \frac{Y-k}{(X+k)+\lambda(Y-X-2k)}$$

The sufficient condition can therefore be restated as:

$$\frac{\partial f(\theta)}{\partial \theta} \geq \frac{Y-k}{X+k} \ \forall \theta$$

The right-hand side variable is decreasing in k . We therefore have:

$$\frac{Y}{X} > \frac{Y-k}{X+k} \ \forall k$$

This gives us a neater version of the sufficient condition for Integral X to start at a higher value than Integral Y:

$$\frac{\partial f(\theta)}{\partial \theta} \geq \frac{Y}{X} > 1 \ \forall \theta \qquad \textbf{Rate of increase}$$

As an aside we note that this condition is never met with a uniform distribution.

Step 3: Calculating the slopes

The impact of a change in type on each of the integrands is calculated, for $s = Y-k$, $X+k$, as:

$$\frac{\partial}{\partial \theta}[\text{Integrand}] = \frac{\partial W(s,\theta)}{\partial q(\theta)} \frac{\partial f(\theta)}{\partial \theta} - \lambda f(\theta)$$

With $\frac{\partial f(\theta)}{\partial \theta} > 1$, $\frac{\partial W(s,\theta)}{\partial q(\theta)} > 0$, $\lambda \geq 0$ and $f(\theta) \geq 0$ we have an increasing slope if:

$$\begin{aligned} \frac{\partial W(s,\theta)}{\partial q(\theta)} \frac{\partial f(\theta)}{\partial \theta} - \lambda f(\theta) \quad &> \quad 0 \\ \left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) \quad &> \quad \frac{\lambda}{\frac{\partial W(s,\theta)}{\partial q(\theta)}} \geq 0 \end{aligned}$$

This condition says that, if the Integrands are to be increasing, the curvature of the belief function needs to be larger than the ratio of λ to the rate of change in welfare arising from a change in output.

The right-hand side of this equation is:

$$\frac{\lambda}{\frac{\partial W(s,\theta)}{\partial q(\theta)}} = \frac{b\lambda}{(1-2\lambda)s + \lambda(a-b\theta)}$$

Given $\lambda(a - b\theta) > 0$ we have, for $s = Y - k \; , X + k$:

$$\frac{b\lambda}{(1-2\lambda)\,s} > \frac{b\lambda}{(1-2\lambda)\,s + \lambda(a-b\theta)}$$

The condition for the integrands to have an increasing slope can therefore be reduced to:

$$\left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) > \frac{b\lambda}{(1-2\lambda)\,s}$$

As $Y - k > X + k$ we have:

$$\frac{b\lambda}{(1-2\lambda)\,(X+k)} > \frac{b\lambda}{(1-2\lambda)\,(Y-k)}$$

Giving us a revised condition of:

$$\left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) > \frac{b\lambda}{(1-2\lambda)\,(X+k)}$$

Similarly $X + k > X$ and hence:

$$\frac{b\lambda}{(1-2\lambda)\,(X)} > \frac{b\lambda}{(1-2\lambda)\,(X+k)}$$

The condition for a given value of λ is therefore:

$$\left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) > \frac{b\lambda}{(1-2\lambda)\,X}$$

When we account for all values of $\lambda \epsilon \left(0, \frac{1}{2}\right)$ we have:

$$1 > \frac{\lambda}{1-2\lambda} > 0$$

The sufficient condition for the integrands to be increasing then becomes:

$$\left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) > \frac{b}{X} > 0 \qquad \textbf{Curvature}$$

Step 4: Comparing the slopes

Given a convex welfare function we have

Integrand Y > Integrand X for any given θ

As the Integrand X lies below Integrand Y, for any given type, we need to compare the slopes of these functions to ensure that Integral X is greater than Integral Y.

We know that if we have two increasing functions, A and B, where the slope of A is steeper than the slope of B, the Integral of A starts at a higher point than the Integral of B, and the intervals of integration are the same length, then the Integral of A will be greater than the Integral of B.

Under the conditions derived thus far Integral X starts at a higher value than Integral Y, and both integrands are increasing. Integral X will therefore be larger than Integral Y if the slope of Integrand Y over the region $I_1$is flatter than the slope of Integrand X over I_2. This will be true if the functions are convex functions.

The rate of change in the slope, for $s = Y - k$, $X + k$, is:

$$\frac{\partial^2}{\partial\theta^2}[\text{Integrand}] = \frac{\partial W(s,\theta)}{\partial q(\theta)}\frac{\partial^2 f(\theta)}{\partial\theta^2} - 2\lambda\frac{\partial f(\theta)}{\partial\theta}$$

This derivative must therefore be positive for the function to be convex (ie, with a convex increasing function the slope becomes more positive as the type increases).

Given $\frac{\partial f(\theta)}{\partial\theta} > 0$, $\lambda \geq 0$ and $\frac{\partial W(s,\theta)}{\partial q(\theta)} \geq 0$ we have $\frac{\partial^2}{\partial\theta^2}$[Integrand]$> 0$ if:

$$\begin{aligned}\frac{\partial W(s,\theta)}{\partial q(\theta)}\frac{\partial^2 f(\theta)}{\partial\theta^2} &> 2\lambda\frac{\partial f(\theta)}{\partial\theta}\\ \left(\frac{\partial^2 f(\theta)}{\partial\theta^2}\right)\left(\frac{1}{\frac{\partial f(\theta)}{\partial\theta}}\right) &> \frac{2\lambda}{\frac{\partial W(s,\theta)}{\partial q(\theta)}} > 0\end{aligned}$$

The right-hand side of this equation is:

$$\frac{\lambda}{\frac{\partial W(s,\theta)}{\partial q(\theta)}} = \frac{2b\lambda}{(1-2\lambda)\,s + \lambda(a-b\theta)}$$

As before, given $\lambda(a - b\theta) > 0$, we have:

$$\frac{2b\lambda}{(1-2\lambda)\,s} > \frac{2b\lambda}{(1-2\lambda)\,s + \lambda(a-b\theta)}$$

Using a similar approach to Step 3 we find that the sufficient condition for the Integrands to be convex reduces to:

$$\left(\frac{\partial^2 f(\theta)}{\partial\theta^2}\right)\left(\frac{1}{\frac{\partial f(\theta)}{\partial\theta}}\right) > \frac{2b\lambda}{(1-2\lambda)\,X} > 0 \qquad \textbf{Change in curvature}$$

We know that $\frac{\partial f(\theta)}{\partial\theta} > 0$. A necessary, but not sufficient, condition for the above inequality to hold is therefore:

$$\frac{\partial^2 f(\theta)}{\partial\theta^2} > 0 \qquad \textbf{Convexity}$$

That is, the increasing density function must be convex.

Step 5: Comparing the integrals

If the welfare function is convex, the regulator's belief function is increasing and convex, the rate of increase in the belief function is greater than the ratio of the maximum extremes of the choice sets, and the degree of convexity of the belief function is greater than a positive constant, we have shown that:

1. Integral X starts at a higher point than Integral Y;
2. Both Integrands are increasing in type; and
3. the slope of Integrand X in I_2 is steeper than the slope of Integrand Y in I_1.

Combined these conditions allow us to conclude that the change in welfare from a change in k is positive. That is:

$$\left|\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right| < \left|\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right|$$
$$\Rightarrow \Delta > 0$$

Step 6: Conclusion

If the conditions of the Lemma hold an increase in k leads to an increase in welfare **for any given** X, Y, k. The more restriction there is in the two-interval choice set (ie, the lower is k) the lower is the level of welfare. ■

Lemma F.9 *(Lemma 5.2) Suppose locally:*

(i) the regulator's belief function, $f(\theta)$, is increasing;

(ii) the rate of increase is large enough to offset the decreasing rate of change in welfare:

$$\frac{\partial f(\theta)}{\partial \theta} \geq \frac{Y}{X} > 1;$$

(iii) the degree of curvature of the belief function is greater than a positive integer:

$$\left(\frac{\partial f(\theta)}{\partial \theta}\right)\left(\frac{1}{f(\theta)}\right) > \frac{b}{X}; \text{ and}$$

(iv) the welfare function is concave ($\lambda \epsilon \left(\frac{1}{2}, 1\right)$).

Then welfare from a restricted two-interval choice set decreases with the size of the interval of disallowed output levels. That is, the more restriction there is in the two-interval choice set (ie, the lower is k) the lower is the level of welfare.

Proof. This proof is exactly the same as the proof to Lemma F.9 except that Step 4 is no longer required.

With a concave welfare function ($\lambda \geq \frac{1}{2}$) the rate of change in welfare from a change in type is decreasing in the level of output (Lemma F.3). We therefore have:

$$\frac{\partial W(Y-k,\theta)}{\partial q(\theta)} \leq \frac{\partial W(X+k,\theta)}{\partial q(\theta)}$$
$$\Rightarrow \text{Integrand Y} \leq \text{Integrand X } \textbf{for any given } \theta$$

Given this, the sufficient conditions for $\Delta > 0$ (Integral X>Integral Y) are simply.

1. Integral X starts at a higher value than Integral Y.

2. Both Integrands are increasing in type.

Steps 2, 3 and 5 in the proof to Lemma F.9 show that if the regulator's belief function is increasing, and the rate of increase and the curvature of the belief function are sufficiently large, these sufficient conditions will hold.

We then have:

$$\left|\int_{\frac{a-(X+Y)}{b}}^{\frac{a-2(X+k)}{b}} \frac{\partial W(X+k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right| \geq \left|\int_{\frac{a-2(Y-k)}{b}}^{\frac{a-(X+Y)}{b}} \frac{\partial W(Y-k,\theta)}{\partial q(\theta)} f(\theta)d\theta\right|$$
$$\Rightarrow \Delta \geq 0$$

That is, an increase in k leads to an increase in welfare **for any given** X, Y, k. The more restriction there is in the two-interval choice set (ie, the lower is k) the lower is the level of welfare. ■

Appendix G

Explaining the simulations

We use Matlab© to simulate the output choices of a firm under the regulator's delegated contract. The details of the simulations are given below. Section 5.5 describes the underlying assumptions of the example used, and considers the main results arising from the simulations. Each simulation is repeated three times for $\lambda = 0.25, 0.5, 0.75$.

G.1 Simulation details

In each simulation we set the feasible output levels to be on the interval (0,95]. Output changes in incremental steps of 0.25. This gives the firm a total of 371 outputs to choose from.

Each simulation generates 100 random values of the firm's marginal cost from the uniform distribution on the interval [5,95]. We then find the level of profit for each cost-output combination giving us a 371×100 profit matrix (100 columns=cost types; 371 rows=output levels). The calculation used to find the profit level varies by simulation; details are given below.

The programme finds the maximum level of profit for each type by finding the highest value of profit along each column. It also identifies the level of output which generates that type-specific maximum level of profit - ie, the programme finds a matrix element π_{ij} which is the maximum level of profit for cost type j, and the profit-maximising level of output is the output corresponding to row i.

With the firm's type (marginal cost) and the profit-maximising level of output we can calculate the level of welfare for that type. We take the average value of welfare over the 100 types to get an estimate for expected welfare under the contract for that run of the simulation. We repeat each simulation n times (the value of n is specified below for each contract type), and average the expected welfare from each run to get a final estimated value of expected welfare for the contract.

Unconstrained contract

For the contract with no constraints on the firm's output choices we simply run the programme 500 times (ie, n=500) to get an average level of welfare for a firm faced with the unconstrained contract. For every output level profit is calculated as: $\pi = pq - \theta q$.

Single interval contract

For the single interval contract we set the output floor equal to 5. This affects the profit calculations as follows:

- $q \leq 5 \Rightarrow \pi = -100,000$[1]
- $5 < q \Rightarrow \pi = pq - \theta q$

We run the programme 500 times (ie, n=500) to get an average level of welfare for a firm faced with the single interval contract.

Two interval contract

For the two interval contract we set the output floor equal to 5. In addition the programme generates 100 random values for X and Y. The random variables are restricted so that $5 < X < Y < 47.5$. The last restriction (ie, Y less than the maximum monopoly output level) insures that the introduction of the extra interval has an impact on the firm's choice. The simulation is run 100 times for each (X,Y) combination, giving us a total of 10,000 runs.

For this contract profit is calculated as follows:

- $q \leq 5 \Rightarrow \pi = -100,000$
- $5 < q \leq X \Rightarrow \pi = pq - \theta q$
- $X < q \leq Y \Rightarrow \pi = -100,000$
- $Y < q \Rightarrow \pi = pq - \theta q$

We find the level of average welfare for each (X,Y) combination (averaging over the 100 runs). This gives us 100 (X,Y) combinations with their associated expected welfare levels. We then find the (X,Y) combination which yields the highest level of expected welfare and use this as our two interval contract.

[1]The profit level is chosen to be a sufficiently large negative number to ensure that the disallowed output choices never emerge as the profit-maximising choices.

Three interval contract

For the three interval contract we take the welfare-maximising two-interval contract and introduce an additional interval of disallowed output levels. The programme generates values for (W, Z) to form this extra interval. Two different types of simulation are run.

1. *Left-side*: 100 random values for W and Z are randomly generated by the programme so that:
$$5 < W < Z < X < Y < 47.5$$

2. *Right-side*: 100 random values for W and Z are randomly generated by the programme so that:
$$5 < X < Y < W < Z < 47.5$$

Each simulation is run 100 times for each (W,Z) combination, giving us 20,000 simulations in total.

Profit is calculated as follows in the left-side simulations:

- $q \leq 5 \Rightarrow \pi = -100,000$
- $5 < q \leq W \Rightarrow \pi = pq - \theta q$
- $W < q \leq Z \Rightarrow \pi = -100,000$
- $Z < q \leq X \Rightarrow \pi = pq - \theta q$
- $X < q \leq Y \Rightarrow \pi = -100,000$
- $Y < q \Rightarrow \pi = pq - \theta q$

Profit is calculated as follows in the right-side simulations:

- $q \leq 5 \Rightarrow \pi = -100,000$
- $5 < q \leq X \Rightarrow \pi = pq - \theta q$
- $X < q \leq Y \Rightarrow \pi = -100,000$
- $Y < q \leq W \Rightarrow \pi = pq - \theta q$
- $W < q \leq Z \Rightarrow \pi = -100,000$
- $Z < q \Rightarrow \pi = pq - \theta q$

For each simulation type we find the average level of welfare for each (W,Z) combination (averaging over the 100 runs). This gives us 100 left-side (W,Z, welfare) combinations and 100 right-side (W,Z,welfare) combinations. We find the left-side (W,Z) combination and the right-side (W,Z) combination which yields the highest level of expected welfare. We then find the highest welfare level from these two choices to give us our three interval contract.

Four interval contract

For the four interval contract we take the welfare-maximising three interval contract and introduce an additional interval of disallowed output levels. The programme generates values for (U,V) to form this interval. Three different types of simulation are run.

1. *Right-side:* 100 random values for U and V are randomly generated by the programme so that:

$$5 < W < Z < X < Y < U < V < 47.5$$

2. *Left-side:* 100 random values for U and V are randomly generated by the programme so that:

$$5 < U < V < W < Z < X < Y < 47.5$$

3. *Middle:* 100 random values for U and V are randomly generated by the programme so that:

$$5 < W < Z < U < V < X < Y < 47.5$$

Each simulation type is run 100 times (ie, n=100) for each (U,V) combination, giving us 30,000 simulations in total.

Profit is calculated as follows in the right-side simulations:

- $q \leq 5 \Rightarrow \pi = -100,000$
- $5 < q \leq W \Rightarrow \pi = pq - \theta q$
- $W < q \leq Z \leq\Rightarrow \pi = -100,000$
- $Z < q \leq X \Rightarrow \pi = pq - \theta q$
- $X < q \leq Y \Rightarrow \pi = -100,000$

- $Y < q \leq U \Rightarrow \pi = pq - \theta q$
- $U < q \leq V \Rightarrow \pi = -100,000$
- $Y < V \Rightarrow \pi = pq - \theta q$

Profit is calculated as follows in the left-side simulations:

- $q \leq 5 \Rightarrow \pi = -100,000$
- $5 < q \leq U \Rightarrow \pi = pq - \theta q$
- $U < q \leq V \leq\Rightarrow \pi = -100,000$
- $V < q \leq W \Rightarrow \pi = pq - \theta q$
- $W < q \leq Z \Rightarrow \pi = -100,000$
- $Z < q \leq X \Rightarrow \pi = pq - \theta q$
- $X < q \leq Y \Rightarrow \pi = -100,000$
- $Y < q \Rightarrow \pi = pq - \theta q$

Profit is calculated as follows in the middle simulations:

- $q \leq 5 \Rightarrow \pi = -100,000$
- $5 < q \leq W \Rightarrow \pi = pq - \theta q$
- $W < q \leq Z \leq\Rightarrow \pi = -100,000$
- $Z < q \leq U \Rightarrow \pi = pq - \theta q$
- $U < q \leq V \Rightarrow \pi = -100,000$
- $V < q \leq X \Rightarrow \pi = pq - \theta q$
- $X < q \leq Y \Rightarrow \pi = -100,000$
- $Y < q \Rightarrow \pi = pq - \theta q$

For each simulation type we find the average level of welfare for each (U,V) combination (averaging over the 100 runs). This gives us 100 left-side (U,V, expected welfare) combinations, 100 right-side (U,V, expected welfare) combinations, and 100 middle (U,V, expected welfare) combinations. We find the left-side (U,V) combination, right-side (U,V) combination and middle (U,V) combination which yields the highest level of expected welfare. The highest value from amongst these three choices gives the four interval contract.

G.2 Optimal intervals and welfare levels

We present the calculated average welfare levels for each of the contract options discussed above in Tables G.1 to G.3. Comparisons of the welfare levels are discussed in section 5.5.

Table G.1: Simulated Average Welfare Levels when λ is equal to 0.25

No of Intervals	**Floor**	**U**	**V**	**W**	**Z**	**X**	**Y**	**Welfare**
No restriction								592.87
One	5							595.54
Two	5					29.88	36.62	600.95
Three								
Left	5			9.79	29.74	29.88	36.62	612.89
Right	5			37.10	41.26	29.88	36.62	607.96
Four								
Left	5	5.49	7.53	9.79	29.74	29.88	36.62	619.40
Middle	5	29.75	29.78	9.79	29.74	29.88	36.62	630.85
Right	5	42.81	43.54	9.79	29.74	29.88	36.62	617.20

Table G.2: Simulated Average Welfare Levels when λ is equal to 0.5

No of Intervals	**Floor**	**U**	**V**	**W**	**Z**	**X**	**Y**	**Welfare**
No restriction								792.22
One	5							795.52
Two	5					10.37	46.81	843.89
Three								
Left	5			6.32	7.17	10.37	46.81	859.65
Right	5			46.85	46.93	10.37	46.81	859.70
Four								
Left	5	5.18	10.26	46.85	46.93	10.37	46.81	865.75
Middle	5	46.82	46.82	46.85	46.93	10.37	46.81	859.68
Right	5	47.20	47.47	46.85	46.93	10.37	46.81	856.29

Table G.3: Simulated Average Welfare Levels when λ is equal to 0.75

No of Intervals	Floor	U	V	W	Z	X	Y	Welfare
No restriction								992.54
One	5							995.97
Two	5					17.45	26.37	1002.41
Three								
Left	5			10.40	15.69	17.45	26.37	1012.62
Right	5			30.66	31.18	17.45	26.37	1014.10
Four								
Left	5	5.84	6.28	30.66	31.18	17.45	26.37	1019.20
Middle	5	28.40	30.44	30.66	31.18	17.45	26.37	1009.35
Right	5	44.50	46.41	30.66	31.18	17.45	26.37	1009.10